



يوزجيسينج مالايا

**LATIHAN ILMIAH II
(WXET3182)**

JAWI CHARACTER RECOGNITION

DISEDIAKAN OLEH

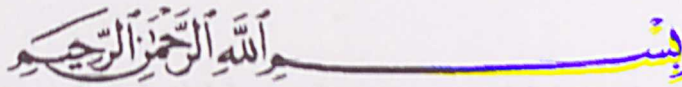
**OMAR BIN ABDUL RAHIM
WET990146**

**IJAZAH SARJANA MUDA TEKNOLOGI MAKLUMAT (MULTIMEDIA)
UNIVERSITI MALAYA, KUALA LUMPUR**

DISEDIAKAN UNTUK

**ENCIK MUHAMMAD NIZAM BIN AYUB (MODURATOR)
ENCIK ZAIDI BIN RAZAK (PENASIHAT)**

PENGHARGAAN



Alhamdulillah, syukur Kehadrat Illahi dan segala puji-pujian kepada Allah S.W.T. kerana dengan limpah kurnia dan hidayatNya dapat saya menyiapkan satu kajian yang bertajuk “Jawi Characters Recognition” bagi memenuhi syarat untuk matapelajaran Projek Ilmiah Tahap II (WXET 3182) dengan sempurna dan lancar disamping dapat menyiapkannya dalam jangka masa yang telah ditetapkan. Sesungguhnya tiada sesuatu kerja yang mudah, namun usaha yang berterusan perlu bagi mencapai apa yang dikehendaki walaupun terpaksa berkorban masa, wang ringgit dan sebagainya. Setiap detik pasti mengundang rintangan-rintangan yang kadang begitu menguji tahap kesabaran, kemahiran dan keimanan kita.

Namun kejayaan yang dicapai ini tidak akan hadir dengan sendirinya tanpa bantuan dan tunjuk ajar dari semua pihak terutamanya daripada penasihat saya yang saya hormati *Encik Zaidi B. Razak* dan juga moderator saya *Encik Muhammad Nizam B. Ayub*. Tidak dilupakan juga kepada semua rakan seperjuangan yang banyak memberi nasihat, tunjuk ajar dan bantuan yang walaupun sedikit sumbangan yang diberi, tetapi ia amat amat bermakna bagi saya. Terima kasih saya ucapkan kepada mereka semua dan juga kepada semua yang terlibat secara langsung atau tidak langsung dalam menjayakan projek ini.

Akhir sekali, saya berharap agar kajian ini akan memberi manfaat kepada semua dan menjadi satu panduan untuk mencapai matlamat yang telah ditetapkan. Sekian, terima kasih.

ABSTRAK

Pengecaman imej digital adalah satu kajian yang telah banyak diterokai oleh ahli-ahli profesional dan para pengkaji demi untuk memudahkan pemahaman terhadap bahasa dan sebagainya. Kepelbagaian bahasa dan tulisan di dunai ini telah menarik ramai pengkaji untuk mengkaji bagaimana komputer boleh mengenali dan memahami tulisan ini tanpa perlu menggunakan tenaga manusia. Kajian-kajian ini perlu bagi keseragaman proses penterjemahan khususnya. Hari demi hari teknologi yang semakin canggih telah digunakan bagi melengkapkan kajian yang sedia ada demi mencapai matlamatnya.

Untuk itu terdetik dihati ini untuk mengkaji bagaimana proses ini berlaku. Untuk itu tulisan bangsa kita iaitu bangsa Melayu telah dijadikan sebagai satu bahan untuk dikaji bagaimana kita ingin memecahkan imej tulisan Jawi kepada aksara-aksara tunggal. Kajian ini penting memandangkan proses penterjemahan yang dilakukan sekarang terhadap menuskrip-menuskrip, kitab-kitab ilmu dan bahan-bahan berkaitan amat sukar memandangkan tulisan Jawi zaman dahulu yang tidak seragam dengan tulisan Jawi Moden yang digunapakai sekarang. Selain itu kekurangan tenaga pakar serta keputusan penterjemahan yang tidak selaras dan masa yang terlalu lama diambil untuk menterjemahkan sesuatu dokumen menjadi sebab utama kepada wujudnya kajian ini.

Bertajukkan "Jawi Character Recognition", kajian ini diharap akan menyelesaikan masalah-masalah yang timbul seperti diatas. Semoga ia akan memberikan yang terbaik kepada semua. Insyallah.....

KANDUNGAN

Penghargaan.....	i
Abstrak.....	ii
Kandungan.....	iii-v
1.0 Pengenalan	
1.1 Pengenalan.....	1-2
1.2 Objektif.....	3
1.3 Keperluan Kepada Kajian.....	4
1.4 Jadual Skedul Kajian (Carta Gant).....	5
2.0 Kajian Literasi	
2.1 Pengenalan Tentang Jawi	6-8
2.2 Sejarah.....	9-10
2.3 Tulisan Jawi Lama Vs Tulisan Jawi Moden.....	11
2.4 Kajian Yang Sedia Ada	
2.4.1 Kajian 1.....	12-14
2.4.2 Kajian 2.....	15-17
2.4.3 Kajian 3.....	18-21
3.0 Methodologi dan Keperluan Kajian	
3.1 Pengenalan	
3.1.1 Pemprosesan Imej.....	22

3.1.2	Perwakilan Imej Digital.....	23
3.1.3	Format Fail Digital Imej.....	24
3.2	Analisis Kajian	
3.2.1	Masalah-masalah.....	25-28
3.3	Perjalanan Proses Pengecaman.....	29
3.3.1	Pendigitalan Data	30
3.3.2	Pemprosesan Data.....	30-31
3.3.3	Penapisan Data.....	31
3.3.4	Proses Pengecaman/Recognition.....	32
3.3.5	Keputusan.....	32
3.4	Perjalanan Proses Pengecaman	
3.4.1	Proses Input Data.....	33-34
3.4.2	Proses Segmentasi (Segmentation).....	34
3.4.3	Proses Pemadanan (Matching).....	35-36
3.5	Spesifikasi Perisian.....	37
4.0	Algoritma Pengecaman Imej	
4.1	Algoritma Pengecaman Imej.....	38
4.2	Kaedah Pengujian	
4.2.1	Kaedah Pertama – Kod Berantai.....	39
4.2.2	Kaedah Kedua – Kaedah Pengecaman Bit.....	40
4.3	Tahap Pengujian	
4.3.1	Tahap Pengujian Pertama (Ujian Terhadap Aksara Asas).....	41-42
4.3.2	Tahap Pengujian Kedua (Ujian Terhadap Aksara Tambahan).....	42

4.4	Algoritma Pengecaman Imej Tulisan Jawi.....	43
4.5	Penerangan	
4.5.1	Bentuk Ruang Skala 16 X 16 Untuk Imej.....	44
4.5.2	Tentukan Posisi Titik Tengah Ruang Yang Dicipta.....	44-45
4.5.3	Imej Dilukis/Satukan Pada Ruang Yang Dicipta.....	46
4.5.4	Bahagian Atas, Bahagian Tengah & Bahagian Bawah	
4.5.4.1	Bahagian Atas.....	47
4.5.4.2	Bahagian Tengah.....	47
4.5.4.3	Bahagian Bawah.....	47
4.5.5	Bezakan Setiap Titik Dengan Imej Yang Sedia.....	48
4.5.6	Output/Tamat.....	48
5.0	Fungsi	
5.1	Tugas-tugas fungsi.....	49-50
5.2	Penggunaan Kaedah Kod Rangkaian.....	51
5.3	Cara Pengujian Ko Berangkaian.....	52-53
5.4	Kemungkinan-Kemungkinan Dalam Kod Rangkaian.....	54
5.5	Proses Pengekoden.....	55-57
6.0	Pengekoden.....	58-68
7.0	Kesimpulan	
	Rujukan	

P A P 1

BAB 1

PENGENALAN

1.1 PENGENALAN

Kajian terhadap penerimaan aksara dan objek-objek atau bentuk-bentuk tulisan tangan oleh sesebuah komputer atau “characters recognition” telah lama dilakukan. Perbezaan bahasa dan tulisan hasil daripada negara, bangsa, agama dan budaya yang berlainan telah menggalakkan ramai pengkaji, pengaturcara dan juruanalisa sistem menyelami bagaimana komputer digunakan untuk mengenali bahasa-bahasa ini.

Kewujudan tulisan Jawi sejak dahulu lagi telah menghasilkan pelbagai karya-karya yang penting dalam perkembangan Malaysia. Sebagai satu bahasa pengantara suatu ketika dahulu, Jawi dianggap penting dalam bidang penulisan. Bukan sahaja buku-buku yang ditulis dalam tulisan Jawi, malah penggunaannya begitu meluas meliputi semua bidang.

Namun semua itu tinggal sejarah. Dizaman serba moden dan canggih ini, tulisan Jawi semakin kurang digunakan. Banyak monumen-monumen, buku-buku ilmu dan artefak-artefak dahulu kala yang ditulis dalam Jawi tetapi sehingga sekarang masih belum terungkai maksud dan apa yang tersirat disebaliknya. Penulisan Jawi pada zaman dahulu hanya menggunakan tulisan tangan dan bukan seperti sekarang yang serba canggih dimana tulisan Jawi telah boleh ditaip menggunakan komputer. Ini menghasilkan bentuk tulisan yang pelbagai dan tidak seragam.

Adakah bahan-bahan sejarah zaman dahulu ini hendak dibiarkan sahaja? Jawapannya tentu tidak. Malah kita tidak jemu-jemu cuba menterjemahkan tulisan-tulisan Jawi zaman dahulu yang terdiri dari pada pelbagai bentuk kepada bahasa Jawi moden yang digunakan sekarang. Tulisan Jawi zaman dahulu tidak serupa dengan apa yang kita ketahui hari ini. Kurangnya perhatian terhadap peninggalan zaman dahulu ini akan

menyebabkan tulisan Jawi zaman dahulu ini akan semakin sukar untuk diterjemahkan ditambah pula penggunaan tulisan Jawi yang semakin berkurangan dan minat terhadapnya yang rendah. Adalah menjadi kerugian kepada generasi sekarang jika bahan-bahan sejarah yang juga merupakan bahan-bahan ilmu suatu ketika dahulu dibiarkan begitu sahaja. Ia perlu diterjemahkan untuk dinikmati oleh generasi zaman sekarang walaupun cara, bentuk dan gaya tulisan tangan yang berbeza dan sukar dibaca oleh generasi kita sekarang ini.

Maka tercetuslah idea untuk saya membuat satu kajian yang mendalam terhadap aksara-aksara Jawi ini. Kerumitan mengecam tulisan Jawi zaman dahulu akan diatasi dalam kajian ini. Kita akan teliti bagaimana kita boleh mengenali setiap input imej tulisan tangan yang telah diimbas dengan mesin pengimbas(scanner). Setiap tulisan Jawi yang telah diinput akan dibaca dan dipecah/diasingkan kepada aksara-aksara tunggal. Kita akan lihat bagaimana proses ini berlaku disamping masalah-masalah rumit yang perlu dikenali sebelum komputer benar-benar boleh mengenali tulisan Jawi.

Semoga dengan adanya kajian ini diharap akan memudahkan lagi proses penterjemahan tulisan Jawi zaman dahulu ini yang ditulis dengan tangan kepada yang mudah kita fahami. Dengan segala kemudahan dan teknologi yang ada dibantu oleh alat-alat dan perisian-perisian yang canggih, diharap kajian ini akan berjaya mencapai matlamatnya.

1.2 OBJEKTIF

Untuk kajian ini beberapa objektif penting telah ditetapkan untuk dicapai. Diakhir kajian ini nanti kita akan sama-sama mencapai objektif berikut :-

1. Objektif utama sistem ini adalah untuk program mengenalpasti tulisan aksara Jawi yang ditulis dengan tangan dengan menggunakan kaedah Kod Berangkai.
2. Membolehkan komputer menerima imej tulisan tangan Jawi yang akan di digitalkan dan akan melalui proses Enhancement dan proses-proses lain sebelum ia dapat dikenali oleh komputer atau program.
3. Memecahkan tulisan Jawi samada dalam bentuk perenggan ayat atau perkataan yang ditulis dengan tangan kepada aksara tunggal.
4. Mengatasi masalah untuk menterjemahkan buku-buku, kitab-kitab atau artifek-artifek lama yang ditulis dengan tulisan tangan dimana ada sesetengahnya menggunakan ejaan Jawi lama disamping tulisan tangan yang mempunyai pelbagaian saiz, bentuk dan cara sehingga kadang sukar untuk dibaca.
5. Mengatasi masalah penterjemahan kitab-kitab dan buku-buku ilmu zaman dahulu yang sudah kabur atau terjejas kerana telah terlalu lama disimpan dengan memperbaiki/mengembalikan imej yang lebih terang dan jelas.
6. Mengatasi masalah kekurangan tenaga pakar untuk proses menterjemahkan kitab-kitab lama.

1.3 KEPERLUAN KEPADA KAJIAN

Kajian ini adalah sesuatu yang penting bagi memudahkan proses penterjemahan buku-buku dan kitab-kitab lama yang ditulis dengan tangan. Kitab-kitab dan buku-buku ini mengandungi ilmu yang tidak ternilai harganya, namun sesetengahnya masih tidak dapat diterjemahkan kerana masalah :-

- Tenaga kerja
- Ketidakfahaman terhadap tulisan Jawi dahulu
- Ketidakeragaman tulisan
- Tulisan yang telah kabur dan kurang jelas untuk dibaca.

Tulisan Jawi pada zaman dahulu adalah menggunakan tulisan Jawi Lama yang menggunakan sistem ejaan yang berbeza dengan yang ada sekarang. Malah tulisan tangan yang berbagai bentuk dan gaya yang sesetengahnya amat sukar difahami. Dengan adanya kajian ini, kita tidak perlu risau lagi akan masalah-masalah diatas, dimana buku tadi hanya perlu discan dengan mesin pengimbas dan seterusnya, dalam bentuk digital ia akan diproses sehingga menghasilkan huruf-huruf tunggal secara automatik tanpa perlu merujuk kepada mana-mana rujukan atau panduan. Ia akan menjimatkan masa, tenaga dan seragam dengan mengelakkan kesilapan dalam proses penterjemahan.

Dengan adanya kajian ini, tulisan Jawi ini akan dapat dipecahkan kepada aksara-aksara tunggal yang akan memudahkan proses penterjemahan. Maka ilmu yang ditulis dengan tulisan Jawi pada zaman dahulu ini dapat dimanfaatkan oleh generasi sekarang dan ia tidak hanya menjadi perhiasan di dalam almari sahaja.

2.1. PENGENALAN TENTANG JAWI

Jawi adalah salah satu bahasa yang ditulis dari kanan ke kiri. Sistem ini tidak mempunyai aksara besar atau aksara kecil yang ada pada perbedaan bentuk kepada huruf-hurufnya apabila ditulis secara berurutan atau berangkai. Terdapat 36 aksara Jawi yang digunakan untuk menulis dan membaca kata-kata Melayu.

Jawi adalah salah satu bahasa yang ditulis dari kanan ke kiri. Sistem ini tidak mempunyai aksara besar atau aksara kecil yang ada pada perbedaan bentuk kepada huruf-hurufnya apabila ditulis secara berurutan atau berangkai. Terdapat 36 aksara Jawi yang digunakan untuk menulis dan membaca kata-kata Melayu.

BAB 2

KAJIAN LITERASI

Bayan 1 : rasiyah ini menunjukkan aksara-aksara Arab yang dalam

2.1 PENGENALAN TENTANG JAWI

Tulisan Jawi ialah sejenis tulisan yang ditulis dari kanan ke kiri. Sistem ini tidak mempunyai aksara besar atau aksara kecil yang ada cuma perubahan bentuk kepada aksara-aksara apabila ditulis secara tunggal atau berangkai. Terdapat 36 aksara Jawi yang digunakan untuk mengeja dan menulis kata-kata Melayu.

Jawi seperti yang sedia dimaklumi, merupakan bentuk penulisan Melayu yang menggunakan aksara-aksara Arab, yang sebahagiannya merupakan penerbitan dari aksara-aksara asal seperti yang terdapat dalam Al-Quran. Sebelum penggunaan aksara-aksara rumi, bentuk penulisan jawi digunakan secara meluas sebagai satu alat sebaran Bahasa Melayu. Namun, setelah dijajah oleh British, pengaruh penggunaan aksara-aksara rumi diserapkan ke dalam penulisan-penulisan Bahasa Melayu. Ianya juga diharapkan dapat menyebarkan Bahasa Melayu itu sendiri, dan diharapkan dapat menjadi lingua franca di benua ini. Rajah 1 dibawah menunjukkan aksara-aksara Arab yang asas.

ح	ح	ج	ث	ت	ب	أ
ص	ش	س	ز	ر	ذ	د
ق	ف	ع	ع	ظ	ط	ض
ي	و	ه	ن	م	ل	ك

Rajah 1 : rajah ini menunjukkan aksara-aksara Arab yang utama.

Manakala Rajah 2 dan 3 pula menunjukkan aksara-aksara Jawi yang mana aksaranya telah bertambah kepada 36 daripada aksara Arab yang asal iaitu 28 aksara semuanya.

ح HA	ج JIM	ث THA	ت TA	ب BA	ا ALIF
ز ZAI	ر RA	ذ DZAL	د DAL	خ KHA	چ CHA
ظ DZO	ط THO	ض DHAD	ص SOD	ش SHIN	س SIN
ق QAF	ڤ PA	ف FA	ڠ NGA	غ GHAIN	ع A'IN
و WAU	ن NUN	م MIM	ل LAM	گ GA	ک KAF
ښ NYA	ي YA	ء HAMZAH	لا LAM-ALIF	ه HA	و VA

Rajah 2 : Carta aksara-aksara Jawi dan sebutannya.

h	j	th	t	b	a
s	z	r	z (dh)	d	kh
a (k')	z	t	d	s	sh
m	l	k	k (q)	f	gh
k'	l	h	v	w, u, o	n
g	p	ch	ny	ng	y, i, e

Rajah 3 : Carta aksara-aksara Jawi dan sebutannya dalam tulisan Rumi.

Walaupun tulisan Jawi ini dipinjam daripada tulisan Arab, masih terdapat kekurangan dari sudut lambang-lambang untuk fonem Melayu. Oleh itu orang-orang Melayu telah meminjam beberapa aksara Arab yang telah difarsikan. Dengan itu bertambahlah jumlah aksara Jawi. Aksara-aksara tambahan ini adalah seperti dalam rajah di bawah (Rajah 4).

	daripada aksara		untuk	C
	daripada aksara		untuk	Ny
	daripada aksara		untuk	P
	daripada aksara		untuk	G
	daripada aksara		untuk	L
	daripada aksara		untuk	H
	daripada aksara		untuk	K'
	daripada aksara		untuk	Ng

Rajah 4 : Aksara-aksara tambahan dalam tulisan Jawi.

2.2 SEJARAH

Tulisan Jawi adalah satu seni penulisan yang telah wujud berabad-abad lamanya di negara kita. Kemunculannya berkait secara langsung dengan kedatangan agama Islam ke Nusantara. Ianya berasal dari tulisan Arab dan merupakan aksara-aksara Arab yang dimasukkan ke dalam sistem penulisan bahasa Melayu.

Tulisan Jawi adalah antara tulisan terawal yang pernah ditemui. Tulisan ini telah berkembang sejak zaman Kerajaan Islam Pasai kemudian ke zaman Kerajaan Melaka, zaman Kerajaan Johor dan juga Aceh pada abad ke-17. Bukti kewujudan tulisan ini di Malaysia adalah dengan terjumpanya Batu Bersurat Terengganu yang bertarikh 702H atau 1303M manakala tulisan Rumi yang paling awal ditemui adalah pada akhir kurun ke-19. Ini menunjukkan tulisan Jawi telah sampai ke negara ini lebih awal berbanding tulisan Rumi.

Berdasarkan kepada tarikh batu bersurat yang dijumpai di Terengganu iaitu tahun 1303 didapati tulisan Jawi telah pun berkembang. Dipercayai tulisan Jawi meresap kedalam kebudayaan orang-orang Melayu bersama-sama dengan kedatangan agama Islam yang dipercayai dibawa oleh pedangang-pedangan Arab di Gugusan Kepulauan Melayu.

Sebelum kedatangan agama Islam, orang-orang Melayu telah mempunyai kebudayaan dan tulisan mereka sendiri yang dipengaruhi oleh unsur-unsur Hindu. Antara contoh tulisan mereka yang terkenal pada masa itu ialah Renchong dan Jawi Kuno atau Kawi.

Tulisan Jawi ialah sejenis tulisan yang dipinjam oleh orang-orang melayu daripada tulisan Arab. Nama Jawi ini berasal daripada bahasa Arab "Jawah". Apabila ditukarkan kepada kata adjektif Jawah menjadi Jawi. Istilah Jawi ini hanya dikenali di Malaysia, Thailand Selatan, Singapura dan Brunei. Di Indonesia tulisan ini dikenali dengan "tulisan Melayu aksara Arab" . Penggunaan tulisan jawi (atau istilah yang semakna dengannya) digunakan di banyak tempat di Asia Tenggara. Tulisan ini bukan sahaja digunakan oleh orang-orang Melayu untuk bahasa Melayu tetapi juga untuk suku-suku bangsa Melayu yang lain seperti suku bangsa Melayu Champa di Indo China, Patani di Thailand Selatan, Jawa, Banjar, Aceh Minangkabau, Bengkulu dan lain-lain.

Walaupun tulisan Jawi ini dipinjam daripada tulisan Arab, masih terdapat kekurangan dari sudut lambang-lambang untuk fonem Melayu. Oleh itu orang-orang Melayu telah meminjam beberapa aksara Arab yang telah difarsikan. Dengan itu bertambahlah jumlah aksara Jawi. Ini telah membentuk tulisan Jawi yang begitu unik yang hanya digunakan di Tanah Melayu secara meluas.

2.3 TULISAN JAWI LAMA VS TULISAN JAWI MODEN

Perbezaan tulisan Jawi Moden dan tulisan Jawi Lama dapat kita lihat dengan jelas, dimana tulisan Jawi Moden lebih mengutamakan pengejaan satu demi satu aksara. Manakala tulisan Jawi Lama pula tidak begitu. Dalam tulisan Jawi Moden aksara “Alif” lebih banyak digunakan kerana setiap aksara rumi akan ditukarkan kepada aksara Jawi tanpa meninggalkan apa-apa aksara.

Dalam kajian ini, penentuan terhadap pemilihan samada tulisan Jawi Moden atau tulisan Jawi Lama tidak wujud kerana kajian ini akan hanya memecahkan tulisan Jawi kepada aksara-aksara tunggal dan tiada percantuman semula terhadap aksara yang telah dipisahkan.

2.4 KAJIAN YANG SEDIA ADA

Terdapat beberapa kajian yang serupa iaitu **pengesanan terhadap aksara** yang melibatkan banyak bahasa. Antaranya ialah bahasa Cina, bahasa Jepun, bahasa Arab dan banyak lagi. Namun masih belum banyak lagi penerokaan terhadap penulisan Jawi ini. Untuk mendapatkan kaedah yang sesuai bagi kajian ini, beberapa kajian yang terdahulu telah dikaji. Penelitian terhadap kaedah pengesanan yang digunakan oleh pengkaji-pengkaji sebelum ini penting bagi melihat kaedah yang sesuai bagi kajian terhadap tulisan Jawi ini, memandangkan bentuk tulisan Jawi adalah satu tulisan yang unik.

Untuk itu beberapa kajian yang berlainan dan menggunakan kaedah atau method yang berlainan telah dipilih.

2.4.1 KAJIAN 1

Tajuk : On The Recognition Of Arabic Characters Using Hough Transform Technique.

Kajian oleh : Mohamed Fakir, M. M. Hassani dan Chuichi Sodeyama.

METHODOLOGI YANG DIGUNAKAN

Kajian ini adalah satu kajian pengesanan terhadap tulisan Arab. Imej tulisan Arab yang telah discan akan diproses untuk meningkatkan mutu imej (*enhancement process*). Kemudian proses segmentasi pula dijalankan terhadap aksara. Kaedah yang digunakan ialah kaedah *Hough Transform*. Proses ini akan ini bertujuan untuk memecahkan aksara-aksara Arab tadi.

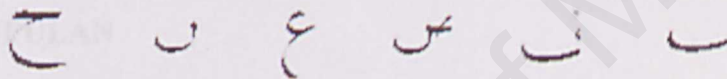
Selepas aksara dipecahkan, kaedah *Dynamic Programming* digunakan bagi proses *matching* iaitu proses bagi mengenali aksara dengan memandangkan aksara yang

telah dipecahkan tadi dengan aksara asas yang telah dimasukkan terlebih dahulu kedalam pengkalan pengetahuan. Proses ini terbahagi kepada dua bahagian iaitu :

- o Badan aksara kan diklasifikasikan menggunakan kaedah *Dynamic Programming*.
- o Selepas itu kaedah topologi pula digunakan unutm mengenali tinggi, lebar dan jumlahnya.

Aksara juga akan dibahagikan kepada 2 bahagian iaitu :

- o Aksara asas (bentuk asas aksara)
- o Aksara tambahan (tambahan kepada bentuk asas iaitu seperti titik)



Aksara asas



Aksara tambahan

HASIL YANG DIPEROLEHI

Hasil daripada kajian yang telah dijalankan purata pengesanan aksara yang tepat mencapai tahap 95%. Dalam proses ini aksara akan dimasukkan dan jika aksara tadi tidak dapat dikenali, maka ia dianggap gagal dan akan ditolak. Peratus yang berjaya adalah tinggi dan ini adalah amat memuaskan.

Terdapat beberapa punca kepada kegagalan pengesanan terhadap aksara Arab ini iaitu. Tulisan yang sukar dibaca sehinggakan aksara sukar dikenali dan jika dikenali pun ia akan mengecam kepada aksara yang lain iaitu menghasilkan keputusan yang tidak tepat.

KESIMPULAN

Kajian ini boleh dijadikan panduan memandangkan tulisan Jawi adalah hampir serupa dengan tulisan Arab. Cuma yang membezakannya ialah dari segi bilangan aksara yang lebih banyak. Aksara Jawi akan menjadi lebih rumit kerana semua aksara tambahan adalah dicipta dari aksara asas Arab yang menghasilkan bentuk yang serupa tetapi hanya dibezakan sedikit oleh bentuk dan titik yang terdapat padanya.

Oleh itu kajian ini sesuai dijadikan panduan utama terutamanya kaedah-kaedah yang digunakan untuk proses segmentasi dan kaedah pemprosesan imej.

2.4.2 KAJIAN 2

Tajuk : On-line Handwriting Character Recognition Method with Directional Features and Direction-Change Features.

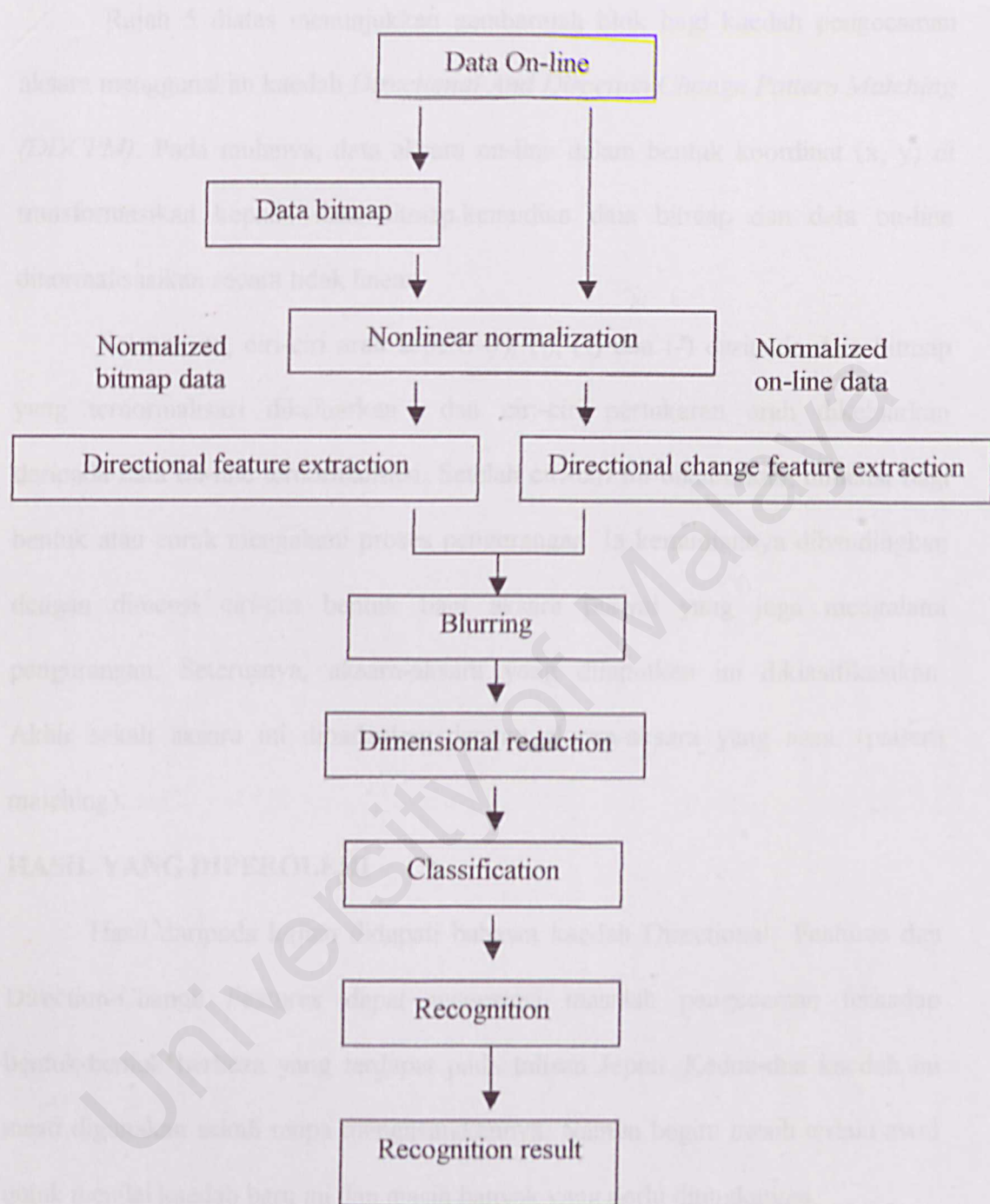
Kajian oleh : Masayoshi Okamoto dan Kazuhiko Yamamoto.

METHODOLOGI YANG DIGUNAKAN

Kajian ini adalah satu kajian terhadap imej tulisan tangan Jepun yang menggunakan kaedah *Directional And Direction Change Pattern Matching (DDCPM)*. Ia menggunakan data on-line dari tulisan yang diinput terus. Kaedah ini diperkenalkan bagi mengatasi masalah kaedah atau bentuk penulisan yang berbagai bentuk iaitu mengikut pergerakan tangan iaitu lengkok, garis melintang, menegak atau mendatar.



Figure 3-1: Flow process of the on-line Directional and Direction Change Pattern Matching (DDCPM)



Rajah 5 : Rajah proses dalam kaedah Directional And Direction Change Pattern Matching (DDCPM)

Rajah 5 diatas menunjukkan gambarajah blok bagi kaedah pengecaman aksara menggunakan kaedah *Directional And Direction Change Pattern Matching (DDCPM)*. Pada mulanya, data aksara on-line dalam bentuk koordinat (x, y) di transformasikan kepada data bitmap. kemudian data bitmap dan data on-line dinormalisasikan secara tidak linear.

Selepas itu, ciri-ciri arah seperti (/), (\), (I) dan (-) daripada data bitmap yang ternormalisasi dikeluarkan dan ciri-ciri pertukaran arah dikeluarkan daripada data on-line ternormalisasi. Setelah ciri-ciri ini dikaburkan, dimensi bagi bentuk atau corak mengalami proses pengurangan. Ia kemudiannya dibandingkan dengan dimensi ciri-ciri bentuk bagi aksara piawai yang juga mengalami pengurangan. Seterusnya, aksara-aksara yang diinputkan ini diklasifikasikan. Akhir sekali aksara ini dipadankan dengan aksara-aksara yang asas. (pattern matching).

HASIL YANG DIPEROLEHI

Hasil daripada kajian didapati bahawa kaedah *Directional Features* dan *Direction-Change Features* dapat mengatasi masalah pengecaman terhadap bentuk-bentuk berbeza yang terdapat pada tulisan Jepun. Kedua-dua kaedah ini mesti digunakan sekali tanpa mengasingkannya. Namun begitu masih terlalu awal untuk menilai kaedah baru ini dan masih banyak yang perlu ditingkatkan.

KESIMPULAN

Kaedah ini masih baru dan belum digunakan secara meluas. Namun hasil yang diberi boleh dikatakan memuaskan. Oleh itu kaedah ini masih boleh diperbaiki.

2.4.3 KAJIAN 3

Tajuk : A Handwriting Character Recognition System Using Directional Element Features and Asymmetric Mahalanobis Distance.

Kajian oleh : Nei Kato, Shin'ichiro Omachi, Hirotomo Aso dan Yoshiaki Nemoto.

METHODOLOGI YANG DIGUNAKAN

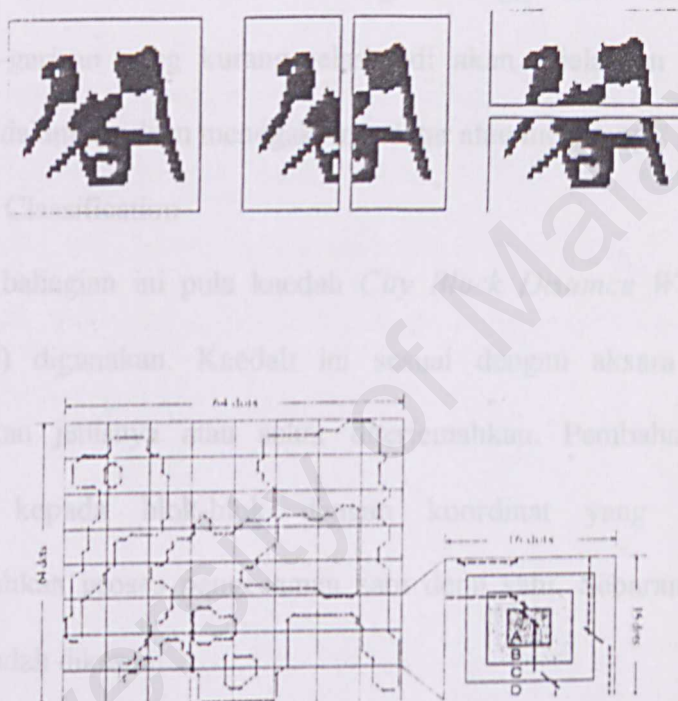
Kajian ini adalah satu kajian pengecaman terhadap tulisan China yang mempunyai 2,965 aksara (*Kanji*) dan Jepun yang mempunyai 71 aksara (*Kana*). Imej yang diperolehi akan menjalani proses peningkatan mutu dengan menggunakan kaedah yang diberi nama *Transformation Based On Partial Inclination* (TPID) bagi membuang segala imej-imej yang tidak diperlukan (*noise*). Hanya tulisan yang sebenar sahaja akan diambil untuk proses pengecaman.

Seterusnya dalam proses pengecaman pula kaedah *City Block Distance With Deviation* (CBDD) dan *Asymmetric Mahalanobis Distance* (AMD) pula digunakan bagi pengecaman tahap tinggi (sukar) dan pengecaman tahap rendah (mudah). Dalam kajian ini, terdapat 4 langkah utama iaitu :

- Preprocessing (Pra-Pemprosesan)

Proses ini akan mengasingkan semua gangguan-gangguan imej serta semua bentuk noise (image distortion) sebanyak yang mungkin dimana kaedah *Transformation Based On Partial Inclination* (TPID) digunakan sebelum proses mencantik dan melicinkan lagi imej tadi dilaksanakan.

Imej aksara akan dipecahkan kepada 2 bahagian samada secara menegak atau melintang. Kemudiannya setiap sudut dan bentuk kecondongan akan dikira. Rajah 6 dibawah menjelaskan prosesnya.



Rajah 6 : Rajah menunjukkan bagaimana proses pemecahan imej dan seterusnya bentuk akhir imej sebelum diproses

- Feature Vector Extraction

Dalam bahagian ini pula kaedah *Directional Element feature* (DEF) digunakan. Kaedah ini akan akan memproses imej telah diproses. Imej yang tinggal hanya garisan halus yang diwakili oleh koordinat-koordinat yang membentuk aksara. Namun akan terdapat juga kesan-kesan kabur yang sukar untuk ditentukan. Dengan menggunakan kaedah *Skeleton* garisan-garisan yang kurang jelas tadi akan dijelaskan tidak kira ia berada dalam keadaan menegak, mendatar atau melintang.

- Rought Classification

Dalam bahagian ini pula kaedah *City Block Distance With Deviation* (CBDD) digunakan. Kaedah ini sesuai dengan aksara yang sukar ditentukan jenisnya atau sukar diterjemahkan. Pembahagian aksara-aksara kepada blok-blok dengan koordinat yang diberi akan memudahkan proses pengecaman satu demi satu. Sebarang perubahan kecil mudah dikesan.

- Fine Classification

Kaedah *Asymmetric Mahalanobis Distance* (AMD) pula digunakan disini. Ia berbeza dengan kaedah diatas dimana setiap sampel asymmetric (bentuk-bentuk) akan ditentukan jenisnya melalui kiraan. Hasil kiraan akan menentukan jenis aksara tersebut.

HASIL YANG DIPEROLEHI

Keputusannya amat memuaskan dan hampir **mencapai tahap** 100% kerana melalui kaedah ini imej dapat dilihat dengan jelas dan ini memudahkan proses distortion iaitu proses mengasingkan gangguan dan kebisingan dari imej yang ada. Selain itu kekaburan juga dapat dikesan dengan mudah.

KESIMPULAN

Ia menghasilkan keputusan yang amat memuaskan. Namun begitu untuk tulisan yang mempunyai banyak lengkok atau ia akan menjadi sukar untuk menggunakan kaedah ini. Tulisan Jawi bukan sahaja mempunyai banyak lengkok malah terdapat titik-titik yang menentukan sesuatu huruf. Maka kaedah ini tidak beberapa sesuai untuk kajian ini.

3.1 PENGENALAN

3.1.1 Pemprosesan Imaj

Imaj adalah satu bentuk gambar dan juga digital adalah imej yang telah diubah menjadi bentuk digital yang kemudian akan digunakan oleh komputer. Imaj akan diproses dengan cara manusia bergantung kepada kebolehan manusia itu sendiri. Kita akan menggunakan imej sepenuhnya dalam kajian ini bagi mendapatkan hasil yang lebih baik dengan tangan. Oleh itu, kita akan mempelajari beberapa proses yang berkaitan dengan pemrosesan imej digital. Berikut adalah beberapa proses yang akan dilakukan dalam kajian ini:

- Pengiraan area yang aktif
- Pengiraan perimeter
- Pengiraan luas

BAB 3

METODOLOGI DAN KEPERLUAN KAJIAN

Dalam bab ini, kita akan mempelajari beberapa konsep yang berkaitan dengan metodologi dan keperluan kajian.

Daftar:

- Pengiraan area
- Pengiraan perimeter
- Pengiraan luas
- Pengiraan Kuantiti
- Pengiraan Imaj

3.1 PENGENALAN

3.1.1 Pemprosesan Imej

Imej adalah satu bentuk gambar dan imej digital adalah imej yang disimpan dalam bentuk digital yang seterusnya akan digunakan oleh komputer. Imej akan dimanipulasikan oleh manusia bergantung kepada kehendak manusia itu sendiri. Kita akan menggunakan imej sepenuhnya dalam kajian ini bagi mendapatkan tulisan Jawi yang ditulis dengan tangan. Oleh itu imej perlu menjalani beberapa proses yang dipanggil pemprosesan imej untuk tujuan itu.

Berikut adalah beberapa proses yang biasa dilakukan keatas sesuatu imej iaitu :

- pengembalian imej yang sebenar
- pembaikpulihan imej
- pemampatan imej

Hasil daripada imej yang telah diproses, outputnya akan digunakan kembali untuk analisis yang seterusnya.

Pemprosesan imej digital seringkali ditafsirkan kepada pemprosesan imej dalam 2 dimensi oleh komputer. Secara am ianya boleh ditunjukkan seperti di bawah:

- Perwakilan Imej
- Penganalisaan Imej
- Pengembalian Imej
- Peningkatan Kualiti Imej
- Pemampatan Imej

3.1.2 Perwakilan Imej Digital

Secara kebiasaannya imej digital diwakili oleh perwakilan $I(r,c)$ di mana r dan c merupakan kedudukan dan piksel tersebut yang menandakan keamatan untuk imej hitam putih. Tetapi untuk imej yang berwarna, perwakilan $I(r,c)$ mewakili fungsi yang berlainan. Berikut adalah beberapa jenis imej :

- **Perduaan**

Perwakilan yang paling senang kerana ianya mempunyai 2 nilai sahaja iaitu putih atau hitam, '0' atau '1'. Kerap kali digunakan dalam aplikasi penglihatan komputer.

- **penskalaan kelabu**

Perwakilan khas yang mengandungi maklumat mengenai keamatan imej. Bilangan bit yang ada menentukan tahap-tahap keamatan sesuatu imej tersebut. Contohnya 8 bit/piksel mengandungi tahap 256 (0-255) keamatan.

- **warna**

Ianya boleh diwakilkan dengan 3-band monochrome imej yang mana setiap band mewakili warna yang berlainan. Sekiranya setiap band mempunyai 8 bit per piksel maka untuk satu piksel dalam perwakilan warna mempunyai 24 bit per piksel. Ini merupakan perwakilan RGB (red, blue, green).

- *pelbagai spektral*

Maklumat yang disimpan menjangkau **darjah penglihatan** manusia.

Contohnya seperti: ultraviolet, X- ray dan radar. **Imej** seperti ini selalunya ditukarkan kepada perwakilan yang lebih bermakna seperti pemetaan kepada RGB

3.1.3 Format Fail Digital Imej

Secara amnya ia boleh dibahagi kepada 2 kategori utama iaitu

- bitmap diwakili oleh permodelan $I(r,c)$
- imej vektor - kaedah yang digunakan untuk mewakili garis, lengkok dan bentuk-bentuk lain yang mana ianya menyimpan titik-titik ataupun ciri-ciri penting sahaja.

Secara amnya format fail yang berjenis bitmap mempunyai 2 bahagian iaitu :

- maklumat kepala yang menyimpan :

- bilangan baris (tinggi)

- bilangan lajur (lebar)

- bilangan jalur

- bilangan bit per piksel

- jenis fail

- data mentah

3.2 ANALISIS KAJIAN

3.2.1 Masalah-masalah

Tulisan Jawi adalah satu tulisan yang unik dan rumit. Sambungan aksara-aksara Jawi untuk membentuk satu perkataan akan merumitkan proses pengecaman. Selain dari ayat yang dibaca dari kiri ke kanan, tulisan Jawi mempunyai dua versi iaitu :

- Tulisan Jawi Lama
- Tulisan Jawi Moden

Perbezaan antara tulisan Jawi Lama dan tulisan Jawi Moden ialah tulisan Jawi Moden menggunakan sepenuhnya sebutan yang ditulis dalam Rumi. Tulisan Rumi ditukar bulat-bulat kepada tulisan Jawi. Setiap aksara rumi akan digantikan dengan aksara Jawi. Ini termasuklah aksara “Alif” yang mewakili aksara Rumi “A”. Dalam tulisan Jawi Lama, aksara “Alif “ digunakan sesekali sahaja dan mengikut tempatnya.

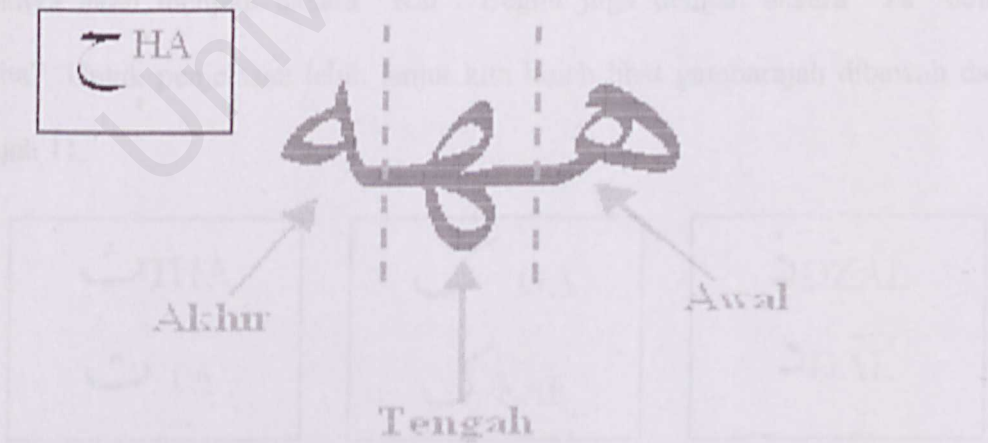
Tetapi semua ini tidak menjadi masalah, kerana apa yang ingin dilakukan disini ialah mengasingkan setiap aksara yang tercantum kepada aksara-aksara tunggal. Masalah pertama yang bakal wujud ialah masalah kedudukan tulisan. Kedudukan tulisan itu penting untuk proses seterusnya. Proses scanning yang kadang terdapat kesilapan akan menjadikan imej tidak berada pada kedudukannya selain daripada imej itu sendiri yang telah semulajadi tidak mendatar. Tulisan hendaklah tidak terbalik atau senget kerana ia akan menyukarkan proses segmentasi. Oleh itu ia perlu diperbetulkan dengan

menggunakan kaedah Hough tranform (HT). Ia akan mengembalikannya kepada keadaan normal.



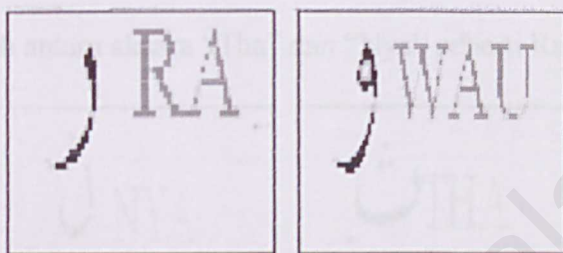
Rajah 8: Rajah contoh keadaan aksara yang tidak sempurna discan atau memang semula jadi keadaannya begini

Selain itu, masalah yang biasa timbul ialah dari segi keadaan tulisan tangan yang kadang sukar dibaca atau silap ejaaan. Ini kerana setiap manusia mempunyaicara tulisannya yang tersendiri. Selain itu tulisan Jawi dengan aksara-aksara sambung yang berlainan bentuk untuk sesuatu aksara amat menyukarkan komputer mengenalinya. Contohnya aksara “Mim” berbeza jika ditulis dihadapan, dibelakang atau ditengah dan penyambungannya dengan aksara-aksara lain akan membentuk aksara “Mim” yang berlainan bentuk dari bentuk asalnya. Berikut adalah rajah bagaimana aksara “Ha” ditulis dihadapan, dibelakang dan ditengah. Contoh boleh dilihat dalam Rajah 9 dibawah.



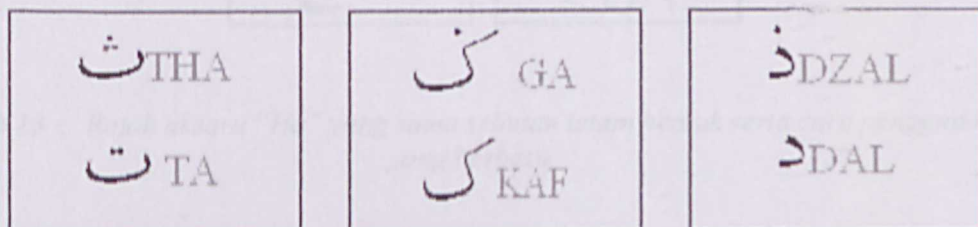
Rajah 9 : Rajah menunjukkan perbezaan bentuk aksara “Ha” jika ditulis di awal, tengah atau diakhir sesuatu perkataan. Rajah kecil menunjukkan aksara asas “Ha”.

Selain masalah diatas masalah tulisan yang tidak sempurna juga menyukarkan proses pengecaman. Bentuk aksara yang hampir serupa antara satu sama lain adalah amat menyulitkan. Contohnya aksara “Wau” dan “Ra” yang kalau ditulis dengan tangan kadang tulisan aksara “Wau” hampir menyerupai aksara “Ra”. Rajah 10 memperlihatkan contoh masalah ini.



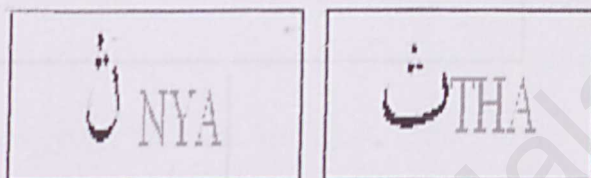
Rajah 10 : Rajah menunjukkan aksara “Ra” dan “Wau” yang mana jika aksara “Wau” tidak ditulis dengan sempurna ia akan menjadi seakan-akan aksara

Seterusnya, masalah yang wujud ialah masalah kesempurnaan tulisan dimana kadang titik untuk sesuatu aksara Jawi itu tidak kelihatan samada kita terlupa atau titik itu tidak dapat dikesan. Ini akan memberi kesan yang besar dalam proses pengecaman aksara. Contohnya aksara “Ga” jika tidak dikesan titiknya akan menjadi aksara “Kaf”. Begitu juga dengan aksara “Ta” dengan “Tha”. Untuk penjelasan lebih lanjut kita boleh lihat gambarajah dibawah dalam Rajah 11.



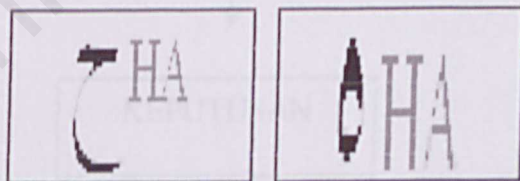
Rajah 11 : Rajah menunjukkan contoh-contoh yang menunjukkan aksara-aksara Jawi yang sama bentuk tetapi berlainan dari segi titik sahaja. Oleh itu titik juga penting bagi menentukan sesuatu aksara. Jika satu titik tidak dapat dikesan maka aksara yang sepatutnya diperolehi akan berubah menjadi aksara yang lain.

Masalah lain ialah aksara yang hampir sama bentuk tetapi berbeza sebutan. Ia hanya dibezakan sedikit sahaja antara satu sama lain. Bezanya terletak pada kelebaran aksara dan ini tentu sekali menyukarkan proses pengecaman. Contoh yang jelas ialah antara aksara “Tha” dan “Nya” seperti Rajah 12 di bawah.



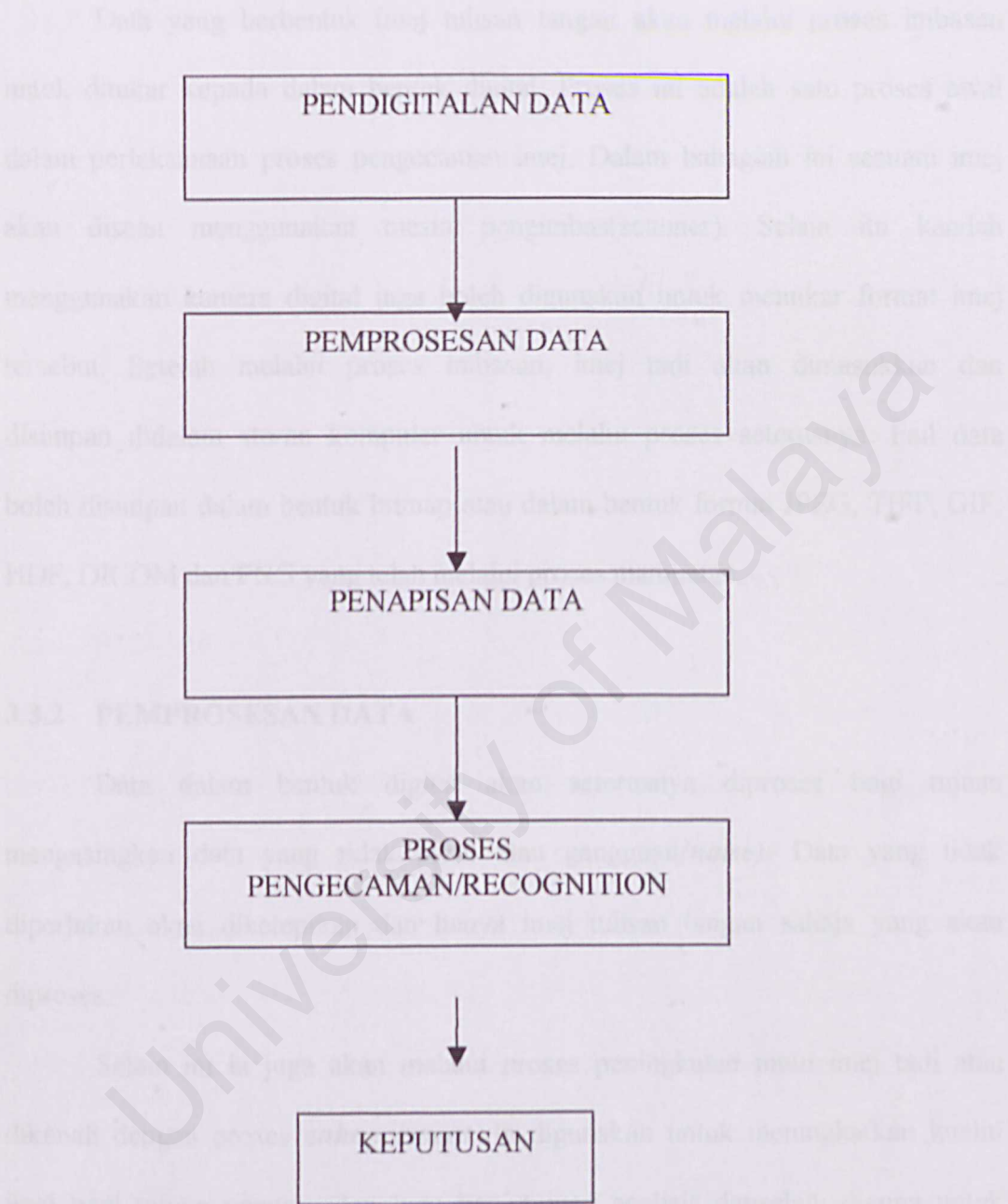
Rajah 12 : Rajah menunjukkan aksara “Tha” dan “Nya” yang hampir serupa

Masalah lain terakhir ialah terdapat aksara yang berlainan bentuk tetapi sebutannya yang sama. Contohnya ialah aksara “Ha” yang sama sebutannya tetapi mempunyai mempunyai bentuk yang berlainan. Penggunaan aksara ini ditentukan berdasar keadaan. Berikut adalah aksara “Ha” yang dimaksudkan (Rajah 13).



Rajah 13 : Rajah aksara “Ha” yang sama sebutan tetapi bentuk serta cara penggunaan yang berbeza.

3.3 PERJALANAN PROSES PENGECAMAN TULISAN JAWI



Rajah 14 : Rajah proses perjalananan pengecaman imej

3.3.1 PENDIGITALAN DATA

Data yang berbentuk imej tulisan tangan akan melalui proses imbasan untuk ditukar kepada dalam bentuk digital. Proses ini adalah satu proses awal dalam pelaksanaan proses pengecaman imej. Dalam bahagian ini sesuatu imej akan discan menggunakan mesin pengimbas(scanner). Selain itu kaedah menggunakan kamera digital juga boleh digunakan untuk menukar format imej tersebut. Setelah melalui proses imbasan, imej tadi akan dimasukkan dan disimpan didalam storan komputer untuk melalui proses seterusnya. Fail data boleh disimpan dalam bentuk bitmap atau dalam bentuk format JPEG, TIFF, GIF, HDF, DICOM dan PNG yang telah melalui proses mampatan.

3.3.2 PEMPROSESAN DATA

Data dalam bentuk digital akan seterusnya diproses bagi tujuan mengasingkan data yang tidak perlu atau gangguan(*noise*). Data yang tidak diperlukan akan diketepikan dan hanya imej tulisan tangan sahaja yang akan diproses.

Selain itu ia juga akan melalui proses peningkatan mutu imej tadi atau dikenali dengan proses *enhancement*. Ia digunakan untuk meningkatkan kualiti imej bagi tujuan paparan dan juga bagi tujuan analisis dan selalu diguna untuk menyelesaikan masalah berkenaan dengan penglihatan komputer. Secara amnya apa yang dilakukan oleh proses peningkatan mutu imej. Proses ini amat penting kerana imej yang terhasil daripada tulisan tangan akan mengandungi pelbagai bentuk kekurangan. Antaranya ialah tulisan yang bercelaru iaitu terdapat getaran

tangan pada tulisan, pertindihan tulisan dan juga imej yang kurang jelas dimana tulisan tersebut mungkin datang dari buku-buku atau artifek-artifek yang lama.

Proses seterusnya ialah proses pernormalan(normalization) iaitu satu proses bagi mengurangkan perbezaan kesan tulisan tangan. Dengan itu imej atau tulisan yang terhampir dan tepat akan diperolehi bagi menentukan jenis aksaranya pula. Ia akan merangkumi saiz, bentuk tulisan dan warna tulisan.

Semua proses diatas akan menghasilkan data imej tulisan yang lebih jelas dan terang dimana garisan tang putus atau kabur akan diterangkan dan kelihatan lebih licin dan jelas.

3.3.3 PENAPISAN DATA

Dalam bahagian ini beberapa proses akan dilakukan. Ia adalah bertujuan untuk mengenalpasti jenis aksara. Oleh itu satu kamus atau maklumat mengenai semua jenis bentuk tulisan tangan untuk setiap aksara asas Jawi akan dimasukkan kedalam database program. Aksara tersebut termasuklah aksara asas dan juga aksara yang mungkin jika ia ditulis dalam keadaan bersambung. Kepelbagaian jenis tulisan akan menentukan jenis aksara yang kita kehendaki.

Seterusnya data yang mengandungi imej yang telah diproses tadi akan dipecahkan kepada beberapa bahagian. Satu perenggan atau ayat akan dipecahkan kepada aksara yang bersambung dan yang tidak bersambung. Ia akan ditentukan melalui ruang antara aksara. Jika terdapat ruang kosong maka pecahan aksara akan berlaku.

3.3.4 PROSES PENGECAMAN/RECOGNITION

Proses ini pula adalah satu proses akhir bagi kajian ini dimana fungsi utama proses ini adalah untuk menentukan jenis huruf yang telah dipecahkan samada ia wujud atau tidak. Untuk itu dengan kaedah pengecaman iaitu membandingkan data didalam *library* atau pengkalan data yang telah kita isikan dengan semua jenis bentuk tulisan tangan untuk setiap aksara asas Jawi dengan imej aksara yang telah diasingkan tadi. Proses matching ini akan memastikan kesahihan proses sebelumnya dimana hasil proses sebelum inilah yang akan menentukan samada program akan dapat mengenali sesuatu aksara atau tidak.

Akhir sekali, program akan berjaya memenuhi matlamatnya iaitu menghasilkan output yang dikehendaki. Matlamat tersebut adalah :

- Memecahkan tulisan Jawi yang bersambung kepada aksara tunggal.
- Mengenalpasti aksara yang mempunyai pelbagai bentuk gaya tulisan yang terhasil daripada tulisan tangan.
- Mengenalpasti jumlah huruf yang wujud.

3.3.5 KEPUTUSAN

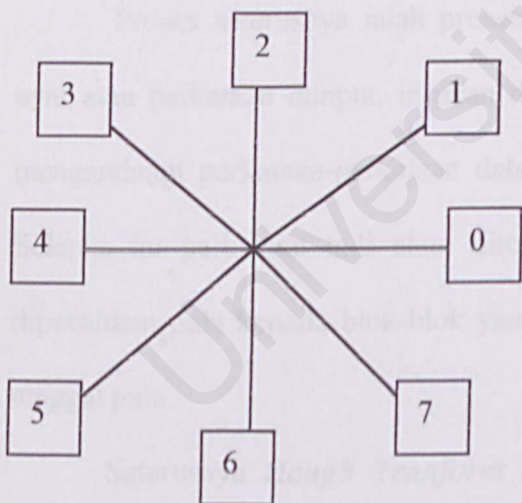
Result adalah keputusan daripada proses pengecaman tadi. Ia akan menghasilkan aksara-aksara tunggal yang telah dipecahkan dari imej tulisan Jawi yang asal. Keputusan akhir ialah keputusan yang akan memberi hasil dari kajian tadi. Keputusan ini akan menentukan samada keseluruhan proses berjaya atau tidak dan ia juga akan menentukan kejayaan kajian ini.

3.4 PERJALANAN PROSES PENGECAMAN

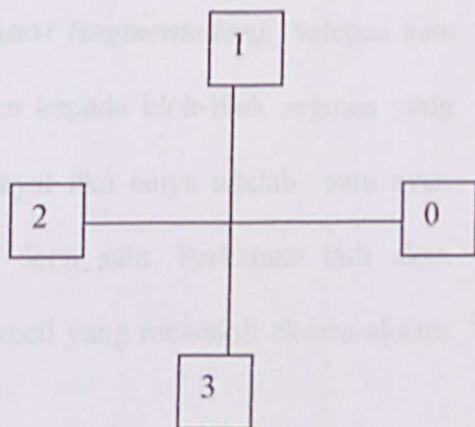
3.4.1 PROSES INPUT DATA

Mula-mula semua aksara-aksara asas Jawi diinputkan kedalam *pengkalan pengetahuan (knowledge base)* sebagai satu kamus dalam proses pengecaman nanti. Kemudian setiap aksara perlu diberikan bentuk-bentuk yang mungkin wujud ketika ditulis dihadapan, ditengah atau dibelakang sesuatu imej perkataan itu. Bentuk-bentuk yang berbeza ini perlu dikenali dan amat penting.

Data-data di atas dimasukkan ke dalam pengkalan pengetahuan dalam bentuk *kod rantaian (chain codes)* dan bukan dalam bentuk imej dalam fail JPEG, GIF, TIFF dan JFIF. Kod rangkaian ialah jujukan nombor-nombor yang mewakili bentuk asas aksara-aksara berdasarkan penakrifan pada *gambarajah arah*.



Rajah 15 : Gambarajah 8 arah kod berantai



Rajah 16 : Gambarajah 4 arah kod berantai

Dari gambarajah 15 dan 16 di atas, kaedah menggunakan gambarajah 8 arah kod berantai lebih popular dan sesuai digunakan memandangkan imej yang

diwakilinya akan kelihatan lebih licin dan sempurna. Setiap bentuk aksara akan diwakili oleh nombor-nombor di atas. Cara bacaannya ialah mengikut putaran jam.

Kaedah penyimpanan data menggunakan perwakilan kod berantai lebih efisien kerana selain ia menjimatkan ruang ia juga memudahkan proses pengecaman aksara atau imej tanpa dipengaruhi oleh mana-mana gangguan dari luar. Ini kerana jika kita membandingkan dua imej pasti sukar kerana kita kena meneliti setiap bahagian pada imej dan perlu pula dipadankan dengan imej yang satu lagi. Ini berbeza jika perwakilan nombor digunakan. Ia akan lebih mudah dan pantas dengan hanya membezakan nombor-nombor yang ada.

3.4.2 PROSES SEGMENTASI (SEGMENTATION)

Proses seterusnya ialah proses *segmentasi (segmentation)*. Selepas satu ayat atau perkataan diinput, ia akan dipecahkan kepada blok-blok segmen yang mengandungi perkataan-perkataan dalam satu ayat jika ianya adalah satu ayat. Selepas itu perkataan tadi akan diteliti satu demi satu. Perkataan tadi akan dipecahkan pula kepada blok-blok yang lebih kecil yang mewakili aksara-aksara tunggal pula.

Seterusnya *Hough Transform (HT)* digunakan sebagai satu mekanisme untuk mengakstrakkan cirian utama bagi menghasilkan aksara yang lebih jelas. Seterusnya proses pengecaman dilakukan dengan membezakan aksara asas dalam *pengkalan pengetahuan* dengan aksara tunggal yang telah dipecahkan.

3.4.3 PROSES PEMADANAN (MATCHING)

Dalam proses ini pula, aksara yang telah dipisahkan dari perkataan dan ayat yang panjang sebelum ini akan di menjalani proses pengecaman (*matching*) untuk mengetahui aksara sebenarnya.

Proses matching aksara ini akan menggunakan kaedah yang dipanggil Dynamic Programming. Untuk proses pengecaman aksara ini, ia dibahagikan kepada dua kumpulan iaitu :

- Aksara asas (bentuk asas aksara)
- Aksara tambahan (tambahan kepada bentuk asas iaitu seperti titik)

ح س ع ص ف ب

Aksara asas

.. .. . ء

Aksara tambahan : Boleh dibahagi 3 iaitu 1) bahagian atas 2) bahagian bawah 3) bahagian tengah

3.5 SPESIFIKASI PERISIAN

Kaedah ini digunakan kerana ia akan menjimatkan storan selain akan mempercepatkan masa pengujian. Jika diteliti, huruf Jawi mempunyai 36 huruf semuanya. Jika kita melalui ujian pengecaman kita perlu mengujinya setiap satu. Oleh itu kaedah memecahkan aksara kepada aksara asas dan tambahan samada diatas, ditengah atau dibawah akan memudahkan dan mempercepatkan proses pengujian.

Ini kerana kebanyakan aksara Jawi mempunyai bentuk aksara(aksara asas) yang sama dan ia cuma dibezakan oleh titik satu, dua dan tiga atau hamzah yang terletak samada diatas, ditengah atau dibawah. Mengikut kajian kita hanya perlu menginputkan 19 bentuk aksara asas sahaja bersama 4 bentuk aksara tambahan.

3.5 SPESIFIKASI PERISIAN

MATHLAB 6.0

Perisian ini dipilih memandangkan ia amat sesuai digunakan untuk proses pengecaman imej. Ia boleh menjelaskan imej dan memperbaikinya disamping dapat digunakan seterusnya untuk tujuan pemprosesan imej tersebut.

- Pentium, Pentium Pro, Pentium II, Pentium III, Pentium IV, atau AMD Athlon based personal computer
- Microsoft Windows 95, Windows 98 (original dan Second Edition), Windows Millennium Edition (ME), Windows NT 4.0 (with Service Pack 5 untuk Y2K compliancy atau Service Pack 6a), atau Windows 2000
- CD-ROM drive (untuk proses installation)
- 64 MB RAM minimum, 128 MB RAM
- 8-bit graphics adapter dan display (untuk 256 simultaneous colors)

Berikut adalah program untuk menulis aturcara bagi proses pengecaman dan pepadanan :

- Microsoft Visual C/C++ version 5.0 atau 6.0
- Borland C/C++ versi 5.0 atau 5.02
- Borland C++Builder versi 3.0, 4.0, atau 5.0

BAB 4

ALGORITMA PENGECAMAN IMEJ

4.1 ALGORITMA PENGECAMAN IMEJ

Algoritma penyelesaian program pengecaman imej tulisan jawi ini akan menumpukan khusus terhadap fungsi mengenali imej dan jenis imej tersebut. Kedua-dua fungsi ini akan bertindak sebagai fungsi yang paling penting dalam menentukan kejayaan kajian ini. Setelah tulisan Jawi yang terdiri daripada ayat/perenggan dipecahkan sehingga menjadi aksara tunggal, maka proses akhirnya adalah dengan cara memecahkan aksara Jawi tunggal tersebut kepada 3 bahagian iaitu :-

- ❑ Bahagian atas (aksara tambahan)
- ❑ Bahagian tengah (aksara asas & aksara tambahan)
- ❑ Bahagian bawah (aksara tambahan)

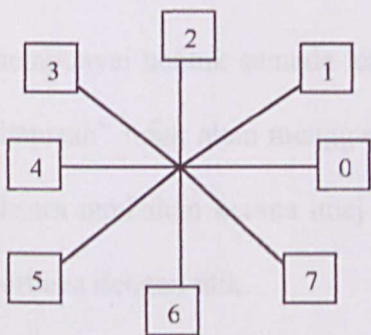
(Aksara tambahan merangkumi titik satu, dua atau tiga serta tanda “hamzah”)

Melalui fungsi ini juga, imej tulisan tangan akan di bezakan dengan imej yang terdapat di dalam database program. Fungsi **jenisimej.m** pula adalah bahagian yang akan memaparkan imej yang telah diproses dan menentukan jenis aksara tersebut.

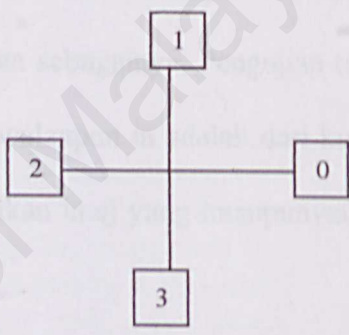
4.2 KAEDAH PENGUJIAN

4.2.1 KAEDAH PERTAMA – KOD BERANTAI

Kod rangkaian ialah jujukan nombor-nombor yang mewakili bentuk asas aksara-aksara berdasarkan penakrifan pada *gambarajah arah*. Setelah diteliti, saya membuat keputusan untuk menggunakan kod rangkaian sebagai satu kaedah membaca/meneliti sesuatu imej aksara. Keadah ini amat sesuai kerana keadaan fizikal tulisan Jawi yang terdapat banyak lengkung, baris dan titik.



Rajah 17 :
Gambarajah 8 arah kod berantai



Rajah 18 :
Gambarajah 4 arah kod berantai

Dari gambarajah 15 dan 16 di atas, kaedah menggunakan gambarajah 8 arah kod berantai lebih popular dan sesuai digunakan memandangkan imej yang diwakilinya akan kelihatan lebih licin dan sempurna.

Setiap bentuk aksara akan diwakili oleh nombor perwakilan kod berantai. Oleh itu setiap nombor akan mewakili bentuk samada menegak, melengkung atau mendatar samada kekiri atau kekanan. Cara bacaannya ialah mengikut putaran jam.

4.2.2 KAEDAH KEDUA – KAEDAH PENGECAMAN BIT

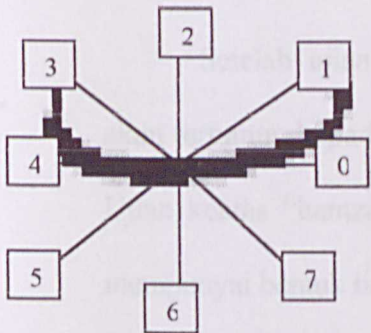
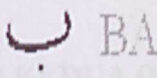
Kaedah pengecaman bit ini adalah satu kaedah bagi mengesan aksara tambahan. Oleh kerana ia bukan merupakan garisan bersambung atau mempunyai bentuk, maka kaedah menentukan samada terdapat titik adalah dengan mengenalpasti setiap koordinat yang mengandungi imej (dot).

Kaedah ini adalah mudah kerana titik tidak merupakan imej yang mempunyai bentuk samada lengkung, garisan dan sebagainya. Pengujian terhadap “hamzah” tidak akan menggunakan kaedah ini walaupun ia adalah dari kumpulan aksara tambahan kerana imej “hamzah” merupakan imej yang mempunyai bentuk berbeza dengan titik.

4.3 TAHAP PENGUJIAN

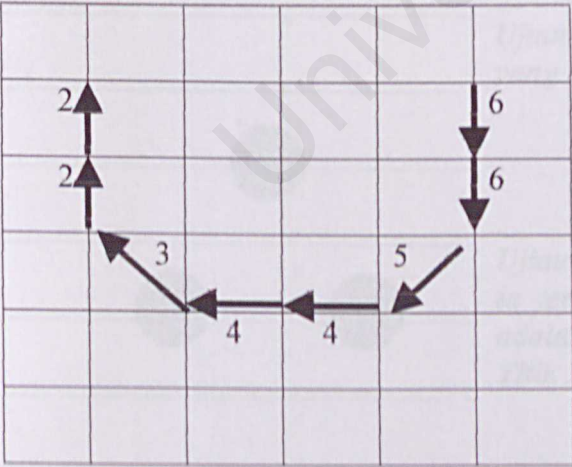
Berikut adalah satu contoh pengecaman aksara yang telah saya tetapkan. Huruf “Ba” akan diambil sebagai contoh.

4.3.1 TAHAP PENGUJIAN PERTAMA (Ujian terhadap aksara asas)



Huruf “Ba” ini telah diasingkan titiknya. Titiknya juga akan diproses seperti ini. Bacaan adalah mengikut titik mana yang mengenai nombor-nombor pada kod berantai itu dan bacaannya adalah mengikut pusingan jam. Bacaannya adalah seperti berikut :
6 6 5 4 4 3 2 2

Hasil ujian keatas titiknya ialah : 2



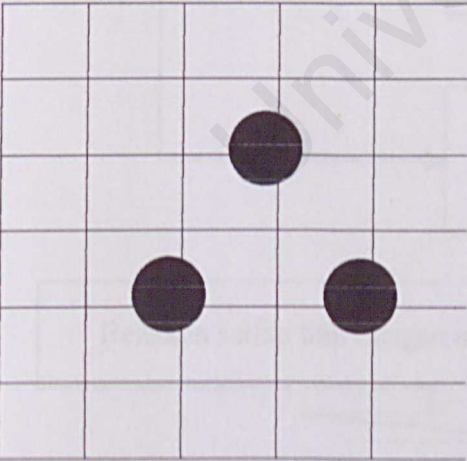
Kaedah sebenar pembacaan mengikut kod rantai 8. Imej huruf “Ba” telah dilakukan pada satu ruang dan → menandakan garisan aksara tersebut.

Rajah 19 : Gambarajah ujian terhadap aksara asas

Proses pengecaman diatas akan membaca atau mengenali aksara diatas mungkin sebagai huruf “Ba”, “Ta” atau “Tha”. Ini kerana didalam database aksara dimasukkan berasingan dengan titik-titiknya dimana aksara-aksara Jawi telah dipecahkan kepada 3 bahagian iaitu bahagian atas, tengah dan bawah. Aksara diatas akan dicam sebagaimana huruf “Ba” hanya apabila ia lengkap/selesai menguji pula jenis titik samada titik tersebut terletak diatas atau bawah.

4.3.2 TAHAP PENGUJIAN KEDUA (Ujian terhadap aksara tambahan)

Setelah ujian pertama selesai terhadap terhadap aksara asas, kini ujian akan tertumpu kepada penentuan aksara tambahan iaitu titik dan tanda “Hamzah”. Ujian keatas “hamzah akan menggunakan kaedah kod berantai 8. Ini kerana ia mempunyai bentuk tidak sebagaimana titik.



Ujian tertumpu kepada penentuan berapa titik yang ada. Tiga kemungkinan yang ada ialah :

- ☐ Satu titik
- ☐ Dua titik
- ☐ Tiga titik

Ujian adalah dengan menentukan koordinat mana ia terletak dan berapa lokasi yang ada. Berikut adalah logik pengecaman titik

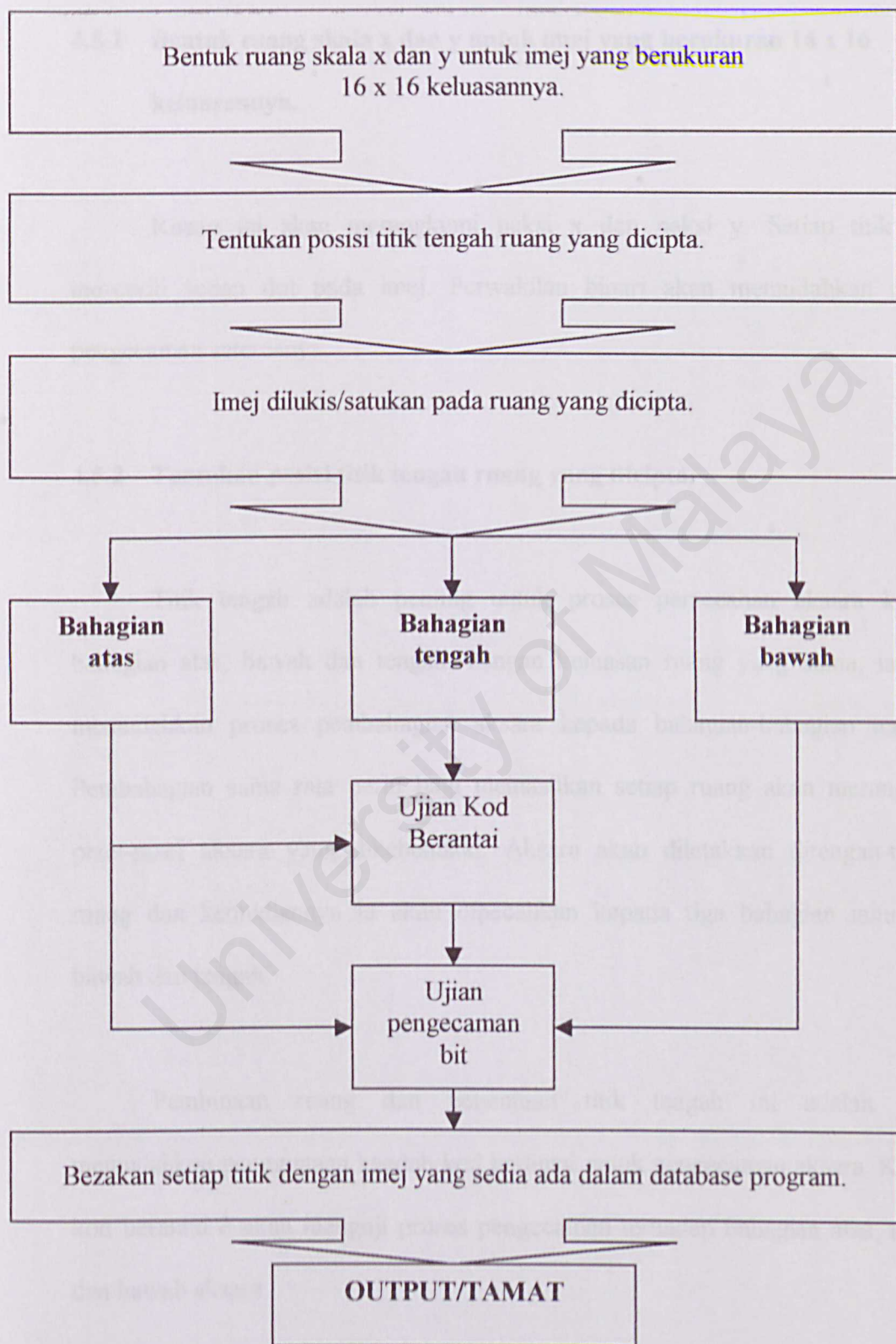
Titik 3 : terdapat 2 koordinat dengan paksi x yang sama serta satu koordinat dengan paksi y berlainan.

Titik 2 : 2 koordinat dengan paksi x yang sama.

Titik 1 : satu koordinat sahaja.

Rajah 20 : Gambarajah ujian terhadap aksara tambahan

4.4 ALGORITMA PENGECAMAN IMEJ TULISAN JAWI



Rajah 21 : Gambarajah perjalanan algoritma pengecaman imej tulisan jawi

4.5 PENERANGAN

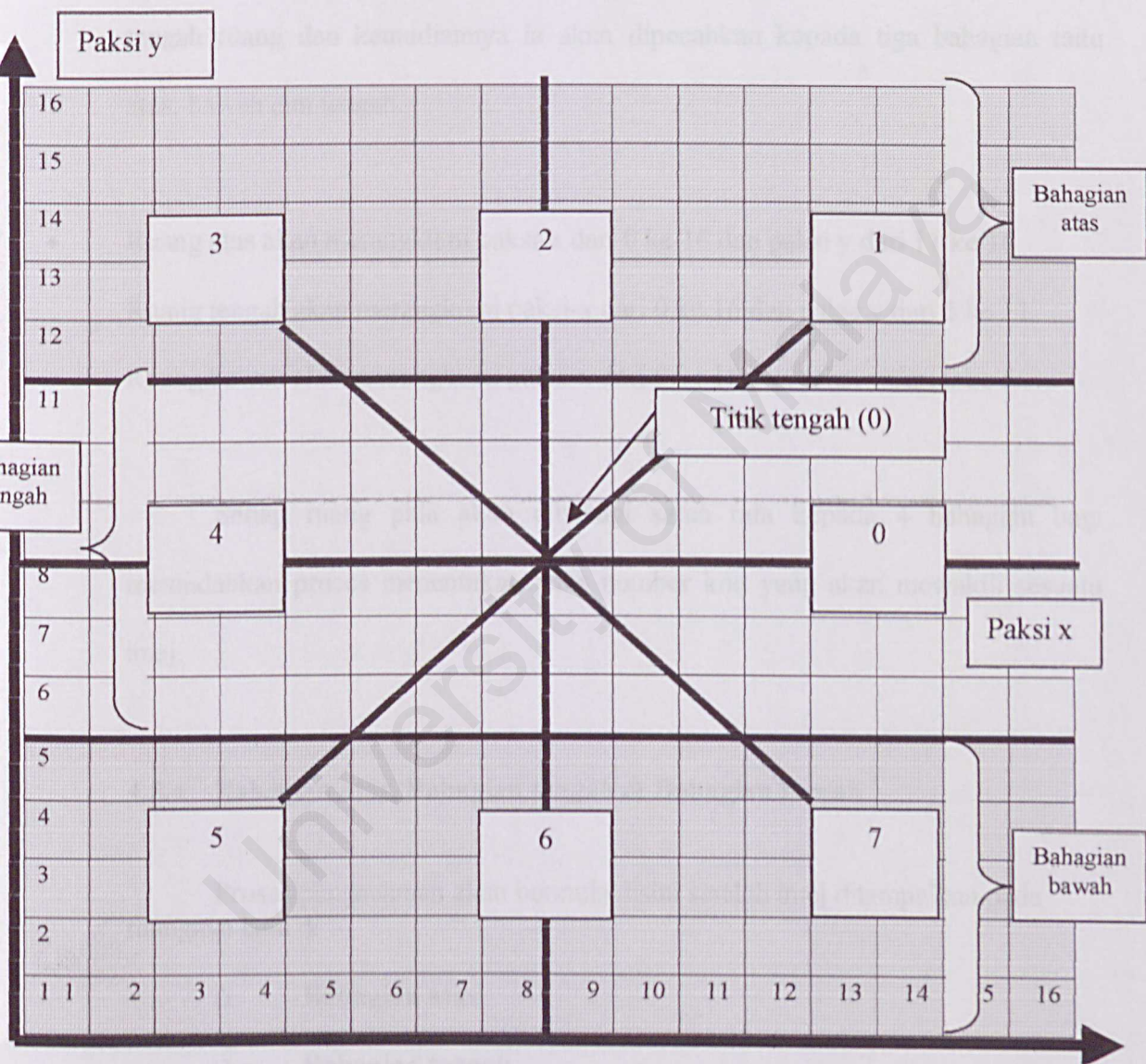
4.5.1 Bentuk ruang skala x dan y untuk imej yang berukuran 16 x 16 keluasannya.

Ruang ini akan merangkumi paksi x dan paksi y. Setiap titik akan mewakili setiap dot pada imej. Perwakilan binari akan memudahkan proses pengecaman seterusnya.

4.5.2 Tentukan posisi titik tengah ruang yang dicipta.

Titik tengah adalah penting untuk proses pemecahan aksara kepada bahagian atas, bawah dan tengah. dengan keluasan ruang yang sama, ia akan memudahkan proses pembahagian aksara kepada bahagian-bahagian tertentu. Pembahagian sama rata perlu bagi memastikan setiap ruang akan merangkumi pixel-pixel aksara yang dikehendaki. Aksara akan diletakkan ditengah-tengah ruang dan kemudiannya ia akan dipecahkan kepada tiga bahagian iaitu atas, bawah dan tengah.

Pembinaan ruang dan penentuan titik tengah ini adalah untuk memudahkan penggunaan kaedah kod berantai untuk pengecaman aksara. Kaedah kod berantai 8 akan menguji proses pengecaman terhadap bahagian atas, tengah dan bawah aksara.



Rajah 22 : Rajah pembahagian ruang 16 x 16 yang dicipta. Ini adalah lakaran gambaran kasar keadaannya.

4.5.3 Imej dilukis/satukan pada ruang yang dicipta.

Setiap bahagian yang telah dicipta akan satu demi satu disatukan/dimasukkan pada ruang dan imej aksara akan diletakkan ditengah-tengah ruang dan kemudiannya ia akan dipecahkan kepada tiga bahagian iaitu atas, bawah dan tengah.

Ruang atas akan merangkumi paksi-x dari 0 ke 16 dan paksi y dari 11 ke 16

Ruang tengah akan merangkumi paksi-x dari 0 ke 16 dan paksi y dari 5 ke 11

Ruang bawah akan merangkumi paksi-x dari 0 ke 16 dan paksi y dari 0 ke 5

Setiap ruang pula akan dibahagi sama rata kepada 4 bahagian bagi memudahkan proses menentukan jenis nombor kod yang akan mewakili sesuatu imej.

4.5.4 Bahagian atas, Bahagian tengah & Bahagian bawah

Proses pengecaman akan bermula disini setelah imej ditampalkan pada ruangnya iaitu di :-

- ☐ **Bahagian atas**
- ☐ **Bahagian tengah**
- ☐ **Bahagian bawah**

4.5.4.1 Bahagian atas

Kaedah pengujian yang digunakan adalah mengikut *kaedah kod berantai 8* bagi penentuan “Hamzah” dan juga *kaedah pengecaman bit* untuk menentukan jenis titik samada satu titik, dua titik atau tiga titik yang terletak diatas aksara asas.

4.5.4.2 Bahagian tengah

Kaedah pengujian yang digunakan adalah mengikut *kaedah kod berantai 8* bagi penentuan aksara asas dan juga *kaedah pengecaman bit* untuk menentukan jenis titik samada satu titik, dua titik atau tiga titik yang terletak ditengah-tengah aksara asas.

4.5.4.3 Bahagian bawah

Kaedah pengujian yang digunakan adalah mengikut *kaedah pengecaman bit* sahaja iaitu untuk menentukan jenis titik samada satu titik, dua titik atau tiga titik yang terletak bawah aksara asas.

4.5.5 Bezakan setiap titik dengan imej yang sedia ada dalam database program.

Setelah itu ia dibezakan dengan imej database yang ada. Untuk itu, tahap 70 peratus adalah kiraan yang dianggap sesuai untuk mengesahkan kesahihan aksara tersebut. Jika ia melebihi tahap kesamaan 70% pada ketiga-tiga bahagian atas, tengah dan bawah, maka ia dapat disahkan aksaranya yang sebenar.

4.5.6 Output/Tamat

Setelah selesai semua proses, ia menukar aksara tulisan tangan tadi menjadi lebih cantik seperti yang ada dalam database. Ini bermakna program telah berjaya mengenali aksara tersebut.

BAB 5

FUNGSI

5.1 Tugas-tugas fungsi

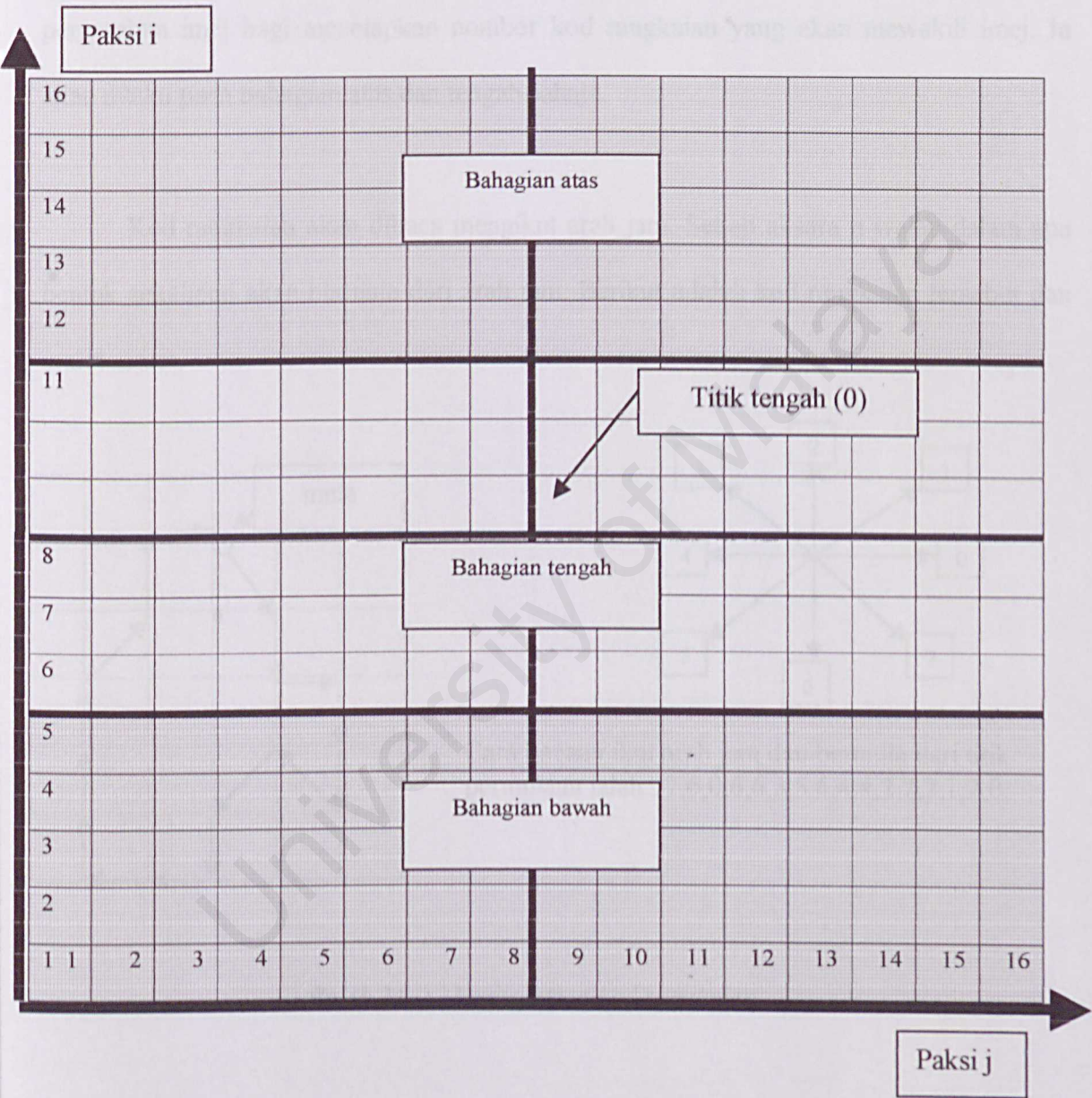
Fungsi ini adalah merupakan fungsi yang mengandungi aturcara penting bagi proses pengecaman aksara Jawi. Antara perkara yang akan dilakukan oleh fungsi ini adalah merangkumi :-

1. Fungsi ini akan memasukkan imej kedalam program. Pengguna boleh memilih direktori mana atau folder mana yang mereka simpan imej bitmap, JPEG, PNG, TIFF dan lain-lain format imej. Imej boleh dicapai dari harddisk, disket, CD-room dan sebagainya.
2. Kemudian imej akan ditetapkan luas pixel dimana satu ruang yang berpaksikan i dan j yang berkeluasan akan dibina.
3. Fungsi ini akan membahagikan ruang tadi kepada 3 bahagian iaitu :-
 - ☐ Bahagian atas (titik atas)
 - ☐ Bahagian tengah (aksara asas)
 - ☐ Bahagian bawah (titik bawah)
4. Tugas fungsi ini juga adalah untuk memastikan ketetapan warna yang seragam. Iaitu hanya satu warna sahaja yang akan diproses iaitu hitam. Untuk itu tulisan imej yang warnanya selain daripada hitam akan ditukar kepada warna hitam.
5. Kemudian proses yang terpenting ialah proses melaksanakan kaedah pengujian kod berantai dan pengecaman bit.

6. Setelah rangkaian nombor mewakili seseatu aksara diperolehi, ia akan dibezakan pula dengan rangkaian kod yang ada di dalam database. Begitu juga dengan kaedah pengecaman bit.
7. Setelah itu barulah disahkan jenis aksara tersebut dan laporan akan dikeluarkan untuk dicetak.

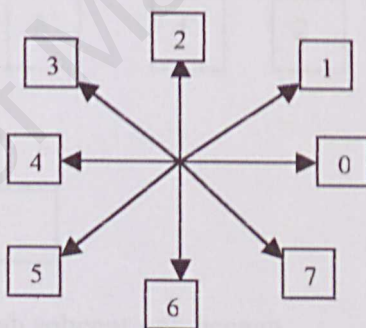
5.2 Penggunaan Kaedah Kod Rangkaian

Mula-mula satu ruang 16 x 16 dibina dan dipecahkan kepada 3 bahagian seperti rajah dibawah.



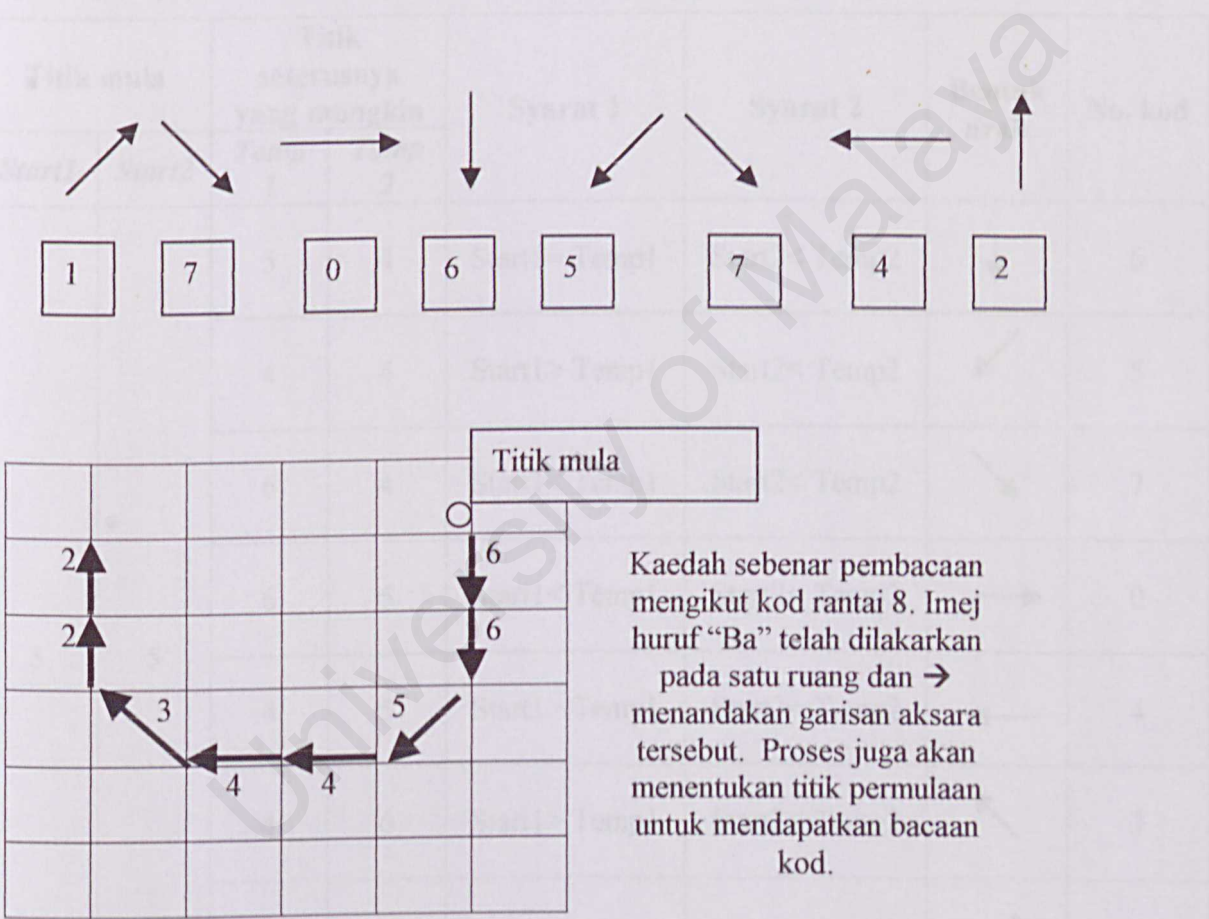
Rajah 23 : Rajah pembahagian ruang 16 x 16 yang dicipta.

2



5.4 Kemungkinan-Kemungkinan Dalam Kod Rangkaian

Oleh kerana kod rangkaian ini dibaca hanya mengikut arah jam, maka akan wujud hanya 8 kemungkinan sahaja bagi setiap kod rangkaian bagi setiap bahagian. Berikut adalah kemungkinan-kemungkinan arah kod rangkaian bagi setiap bahagian. Kemudian fungsi akan mencari titik permulaan bagi sesuatu aksara melalui pembacaan di setiap pixel.



Rajah 25 : Kaedah proses pengecaman menggunakan kod rangkaian

5.4 Kemungkinan-Kemungkinan Dalam Kod Rangkaian

Berikut adalah semua kemungkinan bacaan bagi huruf “Ba” seperti dalam rajah 25 disebelah. Ia akan bermula dari titik koordinat (5,5). Dari titik mula ini akan terdapat 8 kemungkinan pergerakan rangkaian seterusnya. 8 kemungkinan itu ialah :-

Titik mula		Titik seterusnya yang mungkin		Syarat 1	Syarat 2	Bentuk arah	No. kod
Start1	Start2	Temp 1	Temp 2				
5	5	5	4	Start1= Temp1	Start2< Temp2	↓	6
		4	4	Start1> Temp1	Start2< Temp2	↙	5
		6	4	Start1< Temp1	Start2< Temp2	↘	7
		6	5	Start1< Temp1	Start2= Temp2	→	0
		4	5	Start1> Temp1	Start2= Temp2	←	4
		4	6	Start1> Temp1	Start2< Temp2	↖	3
		5	6	Start1= Temp1	Start2< Temp2	↑	2
		6	6	Start1< Temp1	Start2< Temp2	↗	1

5.5 Proses Pengekoden

Seterusnya setelah dilaksanakan dalam program, ia akan berbentuk seperti berikut :-

kod = {0,1,2,3,4,5,6,7}

% Proses menentukan titik mula pembacaan kod rangkaian

for i = 0:16

for j = 0:16

if imx==0

i=start1;

j=start2;

break;

end

end

end

% Proses menentukan titik koordinat dalam imej dan jumlahnya

for i = 0:16

for j = 0:16

if imx==0

i=temp1;

j=temp2;

total+1;

```
if temp1==start1 & temp2==start2
```

```
    if temp2<start2
```

```
        if temp2==start2 - 1
```

```
            kod = 6;
```

```
        else if temp2==start2 + 1
```

```
            kod = 2;
```

```
        end
```

```
    end
```

```
end
```

```
if temp1<start1
```

```
    if temp2==start2
```

```
        kod = 0;
```

```
    else if temp2<start2
```

```
        if temp2==start2 - 1
```

```
            kod = 7;
```

```
        else if temp2==start2 + 1
```

```
            kod = 1;
```

```
        end
```

```
    end
```

```
end
```

```
end
```



```
if temp1>start1
    if temp2==start2
        kod = 4;
    else if temp2<start2
        if temp2==start2 - 1
            kod = 5;
        else if temp2==start2 + 1
            kod = 3;
        end
    end
end
end
end
end
end
end
end
```

BAB 7

PENGEKODAN

```
ion [choice] = testing1;
sgbox('WELLCOME TO MENU UTAMA JAWI HANDWRITTING RECOGNITION');

e = menu(' ==MENU UTAMA== ',' INPUT IMEJ ',' PROSES IMEJ ',' LAPORAN',
ELUAR ');

oice==1

a = msgbox('WELLCOME TO MENU UTAMA JAWI HANDWRITTING RECOGNITION');

Buka fail imej untuk di cam
bmpfile, bmppath] = uigetfile('*.bmp', 'BUKA FAIL IMEJ:');
f (bmppath == 0)
    break;
else
    % Baca fail dan reload imej untuk di cam
    [imx,map]=imread([bmppath,bmpfile]);
    image(imx);
    a = imx;
    colormap(map);
    grid;
    choice = menu(' ==MENU UTAMA== ',' INPUT IMEJ ',' PROSES IMEJ ','
AN ',' KELUAR ');
nd

if choice==2
    % Cari semua jenis warna selain dari hitam
    findwhite=find(imx>0);

    % tukar warna kepada putih
    imx(findwhite)=1;

    % proses bahagikan ruang kepada 3 bahagian

    kod = {'0','1','2','3','4','5','6','7'};
    total = 0;
    % bahagian bawah
    for i = 0:5
        for j = 0:16
            if imx == 0 % ==7 858 -
                total+1;
```



```
        end
    end
end

if total==1
    disp('satu titik');
else if total==2
    disp('dua titik');
else
    disp('tiga titik');
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

[datmat,newmat]=digital;
% bagian tengah
start1=0;
start2=0;
array=[];
total=0;
temp1=0;
temp2=0;
kod{8} = {0,1,2,3,4,5,6,7};
base{3}{64};
% Proses menentukan titik mula pembacaan kod rangkaian
for i = 5:11
    for j = 0:16
        if imx{i}==0
            i=start1;
            j=start2;
            break;
        end
    end
end

% Proses menentukan titik koordinat dalam imej dan jumlahnya
for i = 5:11
    for j = 0:16
        if imx==0
            i=temp1;
            j=temp2;
        end
        if temp1==start1 & temp2==start2
            if temp2<start2
                if temp2==start2 - 1
                    kod{6} :- 59 -
                    array{total} = 6;
                else if temp2==start2 + 1
```

```
kod{2} ;  
array{total} = 2;
```

```
end  
end  
end  
end
```

```
if temp1<start1  
    for j = 1:temp2  
        if temp2==start2  
            kod{0} ;  
            array{total} = 0;  
        else if temp2<start2  
            if temp2==start2 - 1  
                kod{7} ;  
                array{total} = 7;  
            else if temp2==start2 + 1  
                kod{1} ;  
                array{total} = 1;  
            end  
        end  
    end  
end  
end  
end  
end
```

```
if temp1>start1  
    if temp2==start2  
        kod{4} ;  
        array{total} = 4;  
    else if temp2<start2  
        if temp2==start2 - 1  
            kod{5} ;  
            array{total} = 5;  
        else if temp2==start2 + 1  
            kod{3} ;  
            array{total} = 3;  
        end  
    end  
end  
end  
end  
end  
total+1;
```

```
end  
end
```

```
for x = 0:total  
    disp([array]);
```

```
end
```

```
[choice, array] = testing1;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
% bahagian atas
for i = 11:16
    for j = 0:16
        imx;
% uji jenis arah pergerakan imej untuk penetapan nombor kod rangka
ang akan mewakili imej
% untuk proses tentukan kod diruang bahagian atas kiri
        if j<=8 & i>=13.5
            if j~=j & i~=i
                kod = 1;
            else if j==j & i~=i
                kod = 2;
            else if j~=j & i==i
                kod = 0;
            end
        end
    end
end
end

% untuk proses tentukan kod diruang bahagian atas kanan
        if j>=8 & i>=13.5
            if j~=j & i~=i
                kod = 7;
            else if j==j & i~=i
                kod = 6;
            else if j~=j & i==i
                kod = 0;
            end
        end
    end
end
end

% untuk proses tentukan kod diruang bahagian bawah kiri
        if j<=8 & i<=13.5
            if j~=j & i~=i
                kod = 3;
            else if j==j & i~=i
```



```
        kod = 2;

    else if j~=j & i==i
        kod = 4;

    end
    end
    end
    end

% untuk proses tentukan kod diruang bahagian atas kiri
    if j>=8 & i<=13.5
        if j~=j & i~=i
            kod = 5;

        else if j==j & i~=i
            kod = 6;

            else if j~=j & i==i
                kod = 4;

            end
        end
    end
    end
    end
end

end
choice = menu(' ==MENU UTAMA== ',' INPUT IMEJ ',' PROSES IMEJ ',' ↵
AN ',' KELUAR ');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

else if choice==3

    choice = menu(' ==MENU UTAMA== ',' INPUT IMEJ ',' PROSES IMEJ ',' ↵
AN ',' KELUAR ');

    else if choice==4
        a = msgbox('SEKIAN TERIMA KASIH. TAMAT. ');

    end
```

```
on [datmat,newmat]=digital
ses pendigitalan imej
le, bmppath] = uigetfile('*.bmp', 'Buka Fail Imej:');

ppath == 0)
eak

1,y1,button] = ginput(1);

(button == 3) %jika button kanan mouse dipilih
Baca Fail
map]=imread([bmppath,bmpfile]);
age(X);
s image;
ormap(map);

Tentukan skala
msgbox('Tentukan dua titik dengan button mouse sebelah kanan untuk te
ruang yang dikehendaki :');
wait(h)
ton=0;
le button~=3
[x1,y1,button] = ginput(1);

d
ton=0;
le button~=3
[x2,y2,button] = ginput(1);

d
Masukkan nilai maks bagi x dan y
prompt={'Min Xval:', 'Max Xval:', 'Bottom Yval', 'Top Yval'};
le='Sequence Levels';
swer=inputdlg(prompt,title);
x=str2num(char(answer(1,:)));
x=str2num(char(answer(2,:)));
y=str2num(char(answer(3,:)));
y=str2num(char(answer(4,:)));

msgbox('Pilih koordinat : button mouse sebelah kiri untuk baca point
on mouse sebelah kanan untuk -630; ESC untuk quit');
wait(h)
```

```
old on
atmat=[];
utton = 1;
hile button == 1

[xi,yi,button] = ginput(1);

if(button == 3) %jika button kanan mouse dipilih

    plot(datmat(:,1),datmat(:,2),'r')
    h=msgbox('Stop digitalizing! ');
    uiwait(h)
    hold off
    close
elseif(button == 27) %jika ESC dipilih
    h=msgbox(' Stop! Digitalization Interrupted.....');
    uiwait(h)
    close
    clear
    break;
else % jika button kiri dipilih

    if(xi<x1)
        xi = x1;
    end

    if(xi>x2)
        xi = x2;
    end

    if(yi>y2)
        yi = y2;
    end

    if(yi<y1)
        yi = y1;
    end

    % pilih koordinat
    plot(xi,yi,'b+')

    datmat=[datmat; xi yi];

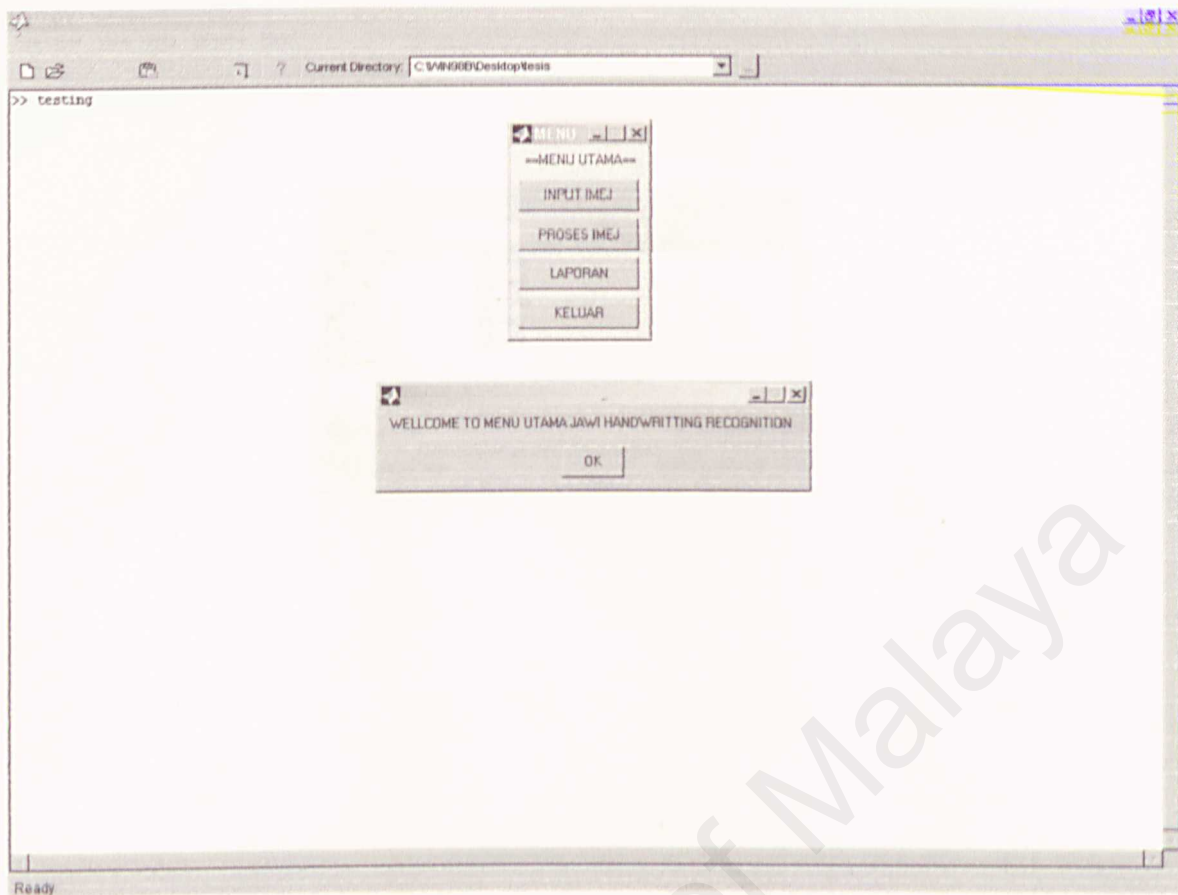
end
datmat=sortrows(datmat,1);
```



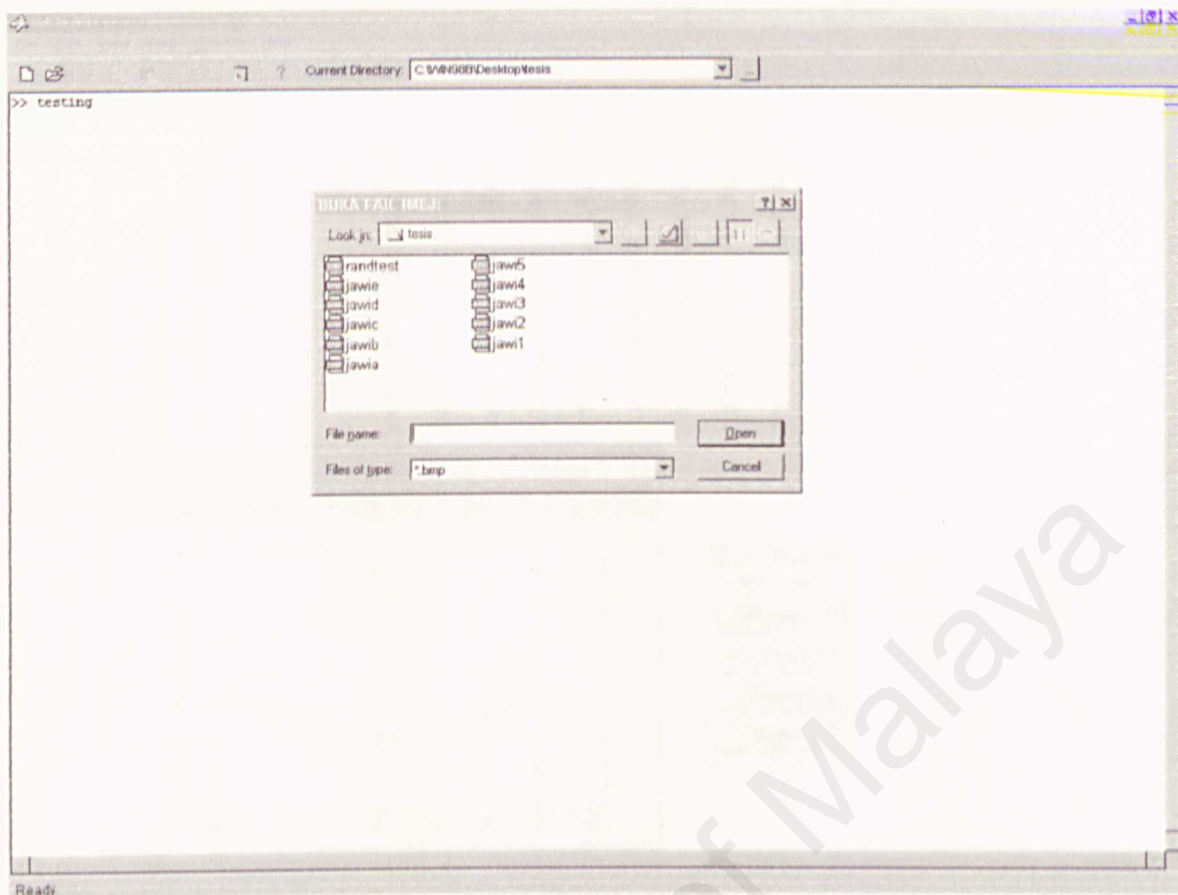
```
scalex=(maxx-minx)/(x2-x1);  
scaley=abs((topy-boty)/(y2-y1));
```

```
lengthy=abs(y2-y1);
```

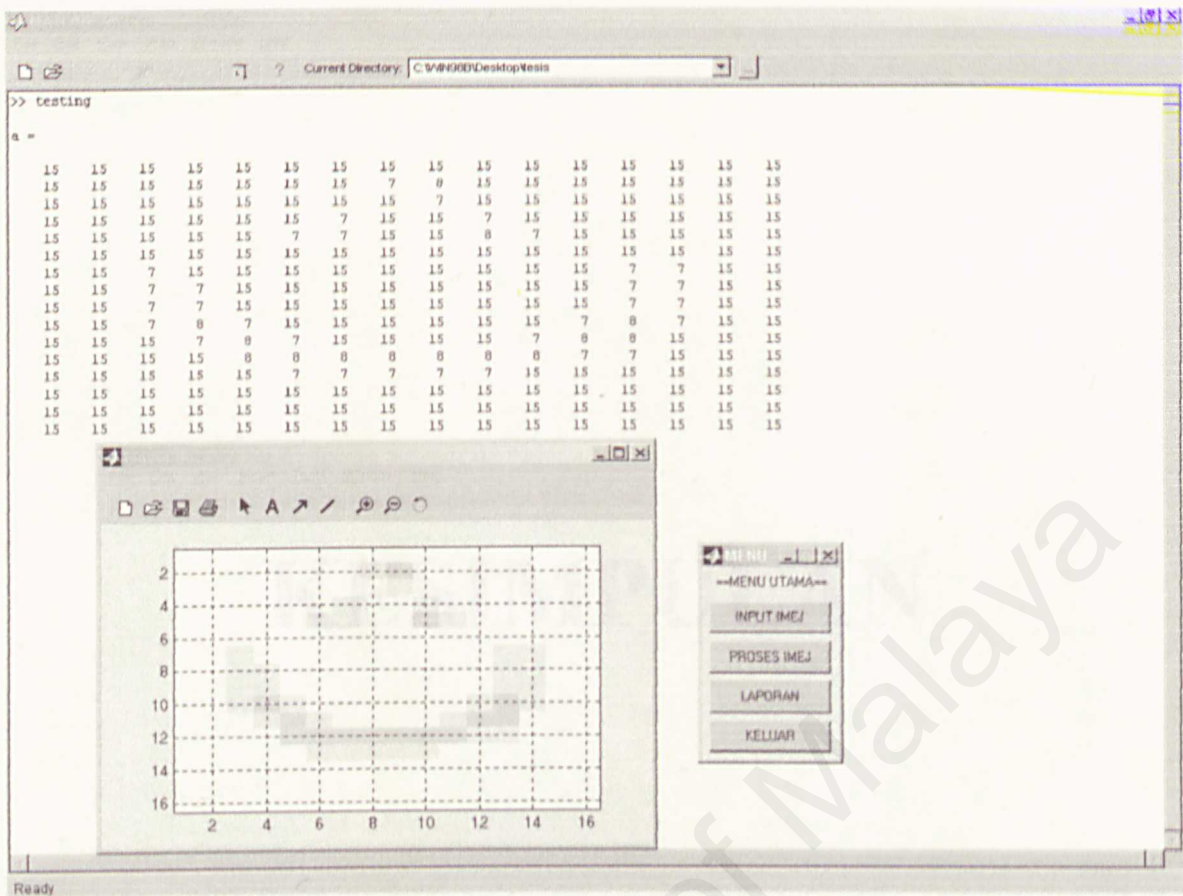
```
if topy>boty  
    newmat=[(((datmat(:,1)-x1).*scalex)+minx),(((lengthy-(datmat(:,2)  
).*scaley)+boty)];  
else  
    newmat=[(((datmat(:,1)-x1).*scalex)+minx),(((datmat(:,2)-y1).*sc  
)+topy)];  
end  
end
```



Antaramuka permulaan



Antaramuka input data



Antaramuka output imej 16x16 bit yang diwakili oleh perwakilan nombor-nombor

KESIMPULAN

KESIMPULAN

KESIMPULAN

Secara keseluruhannya, daripada kajian dan penelitian terhadap kajian yang akan dilakukan untuk matapelajaran Projek Ilmiah Tahap I (WXET 3181) ini, dapatlah disimpulkan bahawa ia telah membuka kepada kajian yang lebih mendalam lagi, lebih menarik dan mencabar bagi menghasilkan keputusan yang terbaik untuk kajian "Jawi Characters Recognition" ini.

Pemilihan kaedah-kaedah yang dianggap terbaik akan melancarkan lagi perjalanan kajian ini kelak. Bermula dengan pemprosesan imej yang telah discan untuk memperelokkannya, seterusnya proses segmentasi untuk memecahkan aksara-aksara dan akhirnya proses memadankan aksara yang diperolehi untuk menentukan jenis aksara., semua ini adalah merupakan proses asas yang diharapkan akan menghasilkan keputusan yang terbaik.

RUJUKAN

- [1] Kamus Inggeris Melayu Dewan, Vetakkan Pertama, Dewan Bahasa dan Pustaka Kuala Lumpur, 1992
- [2] Malaysian Journal Of Computer Science , Volume 13 Number 2, Faculty of Computer Science and Information Technology University of Malaya Kuala Lumpur , June 1997.
- [3] Rafael C. Gonzalez & Richard E. Woods, Digital Image Processing, Addison-wesley Publishing Company, 1993.
- [4] Fu. K. S. & Mui. J.K., "A Survey Of Image Segmentation." Pattern Recognition, vol. 13, 1981.
- [5] J. Hu & H. Yan, "Structural Primitive Extraction and Coding for Handwritten Numeral Recognition", Pattern Recognition, vol 31, 1998.
- [6] A. K. Jain, Fundamentals of Image Processing, Prentice-Hall Inc, Englewood Cliffs 1989.
- [7] David M. Kroenke, Database Processing , Sixth Edition , Prentice-Hall Inc., 1997.
- [8] <http://www.jawinet.8k.com>

[9] <http://www.mathworks.com>

[10] <http://hwr.nici.kun.nl>

[11] IEEE Journal Papers.

University of Malaya